The enzymatic synthesis and bioconjugation of 5'-AzaRNAs

Williamson, David

University of Durham

A Thesis Entitled

The Enzymatic Synthesis and Bioconjugation of 5′–AzaRNAs

Submitted by

David Williamson MChem (Hons)

Department of Chemistry

A Candidate for the Degree of Doctor of Philosophy

2007

Acknowledgements

I would like to thank Dr. David Hodgson for his invaluable support and guidance throughout the course of my research and writing of this thesis. I am also grateful to the EPSRC for funding this research project.

Much of this research would not have been possible without the aid of the departments excellent technical staff, namely: Dr. Alan Kenwright, Mr Ian McKeag and Mrs Catherine Heffernan (NMR); Dr. Mike Jones and Miss Lara Turner (mass spectrometry); Mr Peter Coyne and Mr Malcolm Richardson (glassblowing), Mr Tony Baxter, Mr Jim Lincoln and Mrs Elizabeth Wood (stores) and Dr. Euan Ross (administration).

In addition I would like to thank Dr. Mark Skipsey and Helen Watson for their help and advice in undertaking lab work, as well as their general eagerness to leave work on the promise of a pint.

Finally, I would like to thank my family, friends and girlfriend for providing distractions of in the form of money, utter nonsense and kisses n'stuff.

Cheers.

Memorandum

The work presented within this thesis was carried out at Durham University between October 2003 and September 2006. The thesis is the work of the author, except where acknowledged by reference, and has not been submitted for any other degree. The copyright of this thesis lies solely with the author and no quotation from it should be published without prior written consent and information derived from it should be acknowledged.

This work has been presented at:

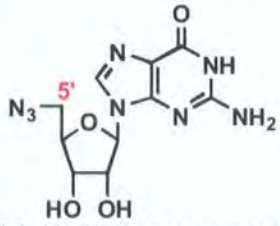
- Durham University Department of Chemistry Final Year Postgraduate Symposium, Durham 2006.
- RSC Organic Reaction Mechanisms Group, Postgraduate Conference, Astra Zeneca, Loughborough 2005.
- RSC Bioorganic Group, Postgraduate meeting, University of Leeds 2004.
- RSC Organic Reaction Mechanisms Group: Reaction Mechanisms VII University College Dublin 2004.

Statement of Copyright

Table of Contents

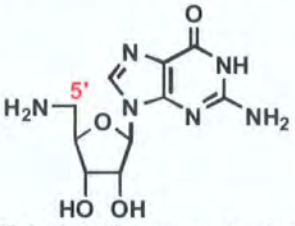
Abstract

As research into the interactions and reactivity of RNA species often requires the bioconjugation of reporter or substrate compounds to the oligonucleotide with high regiospecificity, we envisage uses for RNA molecules modified to incorporate reactive aza groups at the 5′–terminus. By incorporating groups with novel reactivity at specific locations within the RNA molecule, 'handles' are provided to facilitate the specific bioconjugation of the RNA to a wide range of compounds. We achieved the modification through the use guanosine derivatives, modified to feature 5′–aza groups, as initiators in the T7 RNA polymerase mediated synthesis of RNA.



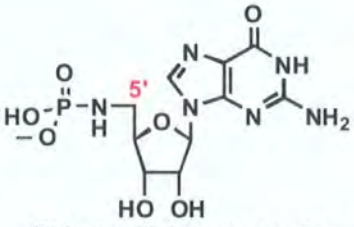
5'-Azido-5'-deoxyguanosine

(Azido–G)



5'-Amino-5'-deoxyguanosine

(Amino–G)



5'-Amino-5'-deoxyguanosine
5'-N-phosphate

(GANP)

The figure above shows the structures of the three initiator species used in our research. The synthesis of GANP from Amino–G was achieved *via* the development of a regiospecific amino–phosphorylation reaction, which takes place in aqueous solution and yields the potentially hydrolytically unstable GANP in high yield (90 %).

Through the use of Amino–G and GANP we were able to improve on literature methods for the synthesis of 5′–AminoRNA, raising the observed level of incorporation from 20 % to a maximum of 88 %[45]. The use of Azido–G enabled us to achieve the unprecedented direct incorporation of an azide group into the 5′–terminus of RNA species (37 %). We successfully demonstrated that the azide group could be reduced to give 5′–AminoRNA or used directly in a copper mediated bioconjugation reaction.

Abbreviations

A	Adenine
Amino–G	5′–amino–5′–deoxyguanosine
Amino–PEG	5′–Amino–hexaethylene glycol guanosine
ATP	Adenosine triphosphate
Azido–G	5′–azido–5′–deoxyguanosine
B	Unspecified Nucleotide base
C	Cytosine
Cac	Cacodylate anion
CHES	2–(cyclohexylamino) ethane sulfonic acid
CTP	Cytosine triphosphate
CuAAC	Copper Catalyzed Azide–Alkyne Cycloaddition
DCM	Dichlorormethane
DEAE	Diethylamine ethylene
DEPC	Diethylpyrocarbonate
DMF	Dimethylformamide
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double Stranded DNA
DTT	Dithiothreitol
EDC	1–Ethyl–3–(3–dimethylaminopropyl) Carbodiimide Hydrochloride
EDTA	Ethylenediamine tetraacetic acid
G	Guanosine
GANP	5′–amino–5′–deoxyguanosine 5′–N–phosphate
GMP	Guanosine monophosphate
GMPS	Guanosine monophosphorothioate
GTP	Guanosine triphosphate
GSMP	5′–deoxy–5′–thioguanosine–5′–monophosphorothioate
HEPES	4–(2–hydroxyethyl)–1–piperazineethane sulfonic acid
Iodo–G	5′–deoxy–5′–iodoguanosine
IP	Image plate
mRNA	Messenger RNA
NHS	N–hydroxysuccinimide
NMP	N–Methyl Pyrrolidinone
NMR	Nuclear Magnetic Resonance
NTP	Nucleotide Triphosphate
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PSL	Photostimulated Luminesence
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SaV	Streptavidin
SELEX	Systematic Evolution of Ligands by Exponential Enrichment

ssDNA	Single Stranded DNA
Succ	Succinate dianion
T	Thymine
T7 RNAP	T7 RNA Polymerase
TCEP	Tris(2–carboxyethyl) Phosphine Hydrochloride
TEAB	Triethylammonium bicarbonate
Thio–PEG	5′–Thiol–hexaethylene glycol guanosine
THF	Tetrahydrofuran
TRIS	2–amino–2–hydroxymethyl–1,3–propanediol
tRNA	Transfer RNA
TTP	Thymidine triphosphate
U	Uracil
UTP	Uridine triphosphate

1.0 Introduction

This first chapter provides a brief introduction to a number of topics relevant to the research we have carried out. All of the chemistry described in our study was carried out using derivatives of nucleic acids. This chapter begins by giving a brief overview of basic nucleic acid molecules, their structure and chemistry, following which we cover the pros and cons of both the chemical synthesis and enzymatic methods of RNA synthesis before giving an overview of bioconjugation sites and methodologies relevant to RNA molecules.

1.1 Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA)

DNA and RNA are two closely related biopolymers at least one of which is present in every cell of every organism. Together they act as a store and method of utilizing inherited genetic information enabling not only their own replication, but also the synthesis of complex proteins such as enzymes, which in turn act to synthesize a variety of other biopolymers. In this way these nucleic acids exert control over the cell in which they are found, dictating everything from its physical structure to its more complex biosynthetic pathways and behaviours.

Of the two species it is DNA that functions as the repository of genetic information in most cells. Given their massive importance within a cell, it would be disadvantageous if DNA strands were regularly exposed to harmful chemical environments that could damage their structure. Hence, it falls to the more common, and more readily reproduced, RNA strands to carry copies of the genetic information held by DNA into the cytoplasm of the cell for use in protein biosynthesis and cell regulation. To undertake this role three distinct forms of RNA have developed:

- Messenger RNA (mRNA) is transcribed directly from DNA before attaching to ribosomes in the cell's cytoplasm where it is translated into protein structures.
- Ribosomal RNA (rRNA) is found complexed with proteins and makes up the ribosome structures that catalyze protein biosynthesis.

- Transfer RNA (tRNA) transports amino acids to the ribosomes where they are linked by peptide bonds in the order dictated by the mRNA sequence.

Many differing forms of each of these RNA types exist to carry out the incredibly complex pathway of protein biosynthesis. However it is beyond the scope of this very brief introduction to go into further detail about the nature of such processes. It is sufficient to say that many forms of RNA exist and though they possess a common basic structure they carry out a variety of tasks within cells.

1.2 The structures of DNA and RNA

The double helix of DNA is arguably the most recognisable chemical structure known to man, though the root of DNA's ability to store information lies in the deoxynucleotides that form this complex structure.

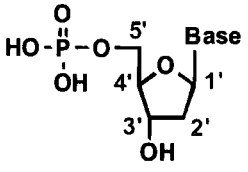


Fig:1.1 A generic deoxynucleotide monophosphate, where **Base** is one of the 4 DNA bases (see Fig:1.2). The conventional numbering pattern for the carbons of ribose is shown.

The four base groups that feature in DNA molecules are cytosine (C), thymine (T), adenine (A), and guanine (G) (Fig:1.2). It is through variation in the order that these four units are arranged throughout the DNA strand that the molecule can store information.

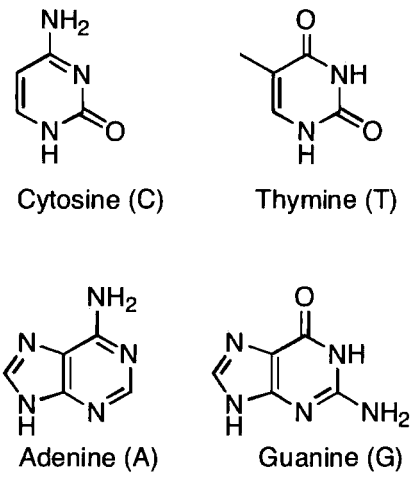


Fig:1.2 The DNA bases, consisting of the two pyrimidine(C and T) and two purine (A and G) derivatives.

The structure of RNA is very similar to that of DNA, though the most obvious difference is that RNA exists in both single and a variety of double stranded structures. That is not to say that ordered structures do not exist within RNA strands, just that those that do are not as well known or stable as the DNA double helix. As with DNA, RNA is made up of smaller nucleic acid units:

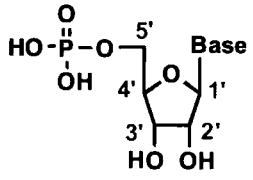


Fig:1.3 A generic ribonucleotide monophosphate, where **Base** is one of the 4 RNA bases (see Fig:1.4). The numbering pattern shown for the carbons of the ribose sub-structure is that used by convention.

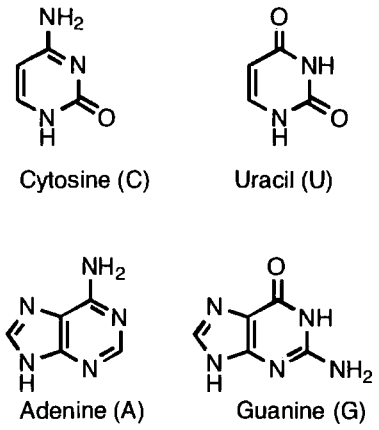


Fig:1.4 The RNA bases, consisting of the two pyrimidine(C and U) and two purine (A and G) derivatives.

Both DNA and RNA strands feature stable sugar–phosphate backbones made up by phosphodiester linkages between the 5′ and 3′ positions of nucleotide monomer. From this backbone the bases protrude outwards, making bases of both RNA and DNA available to take part in Watson–Crick binding, either with a complementary strand (as in DNA) or with a complementary section within the same strand (in RNA).

As Fig:1.1 shows each nucleotide monomer consists of a phosphate group and a 2'deoxyribose sugar, linked to one of four heterocyclic bases. The bases present in DNA are two purine derivatives, adenine and guanine, and two pyrimidine derivatives, cytosine and thymine (Fig:1.2).

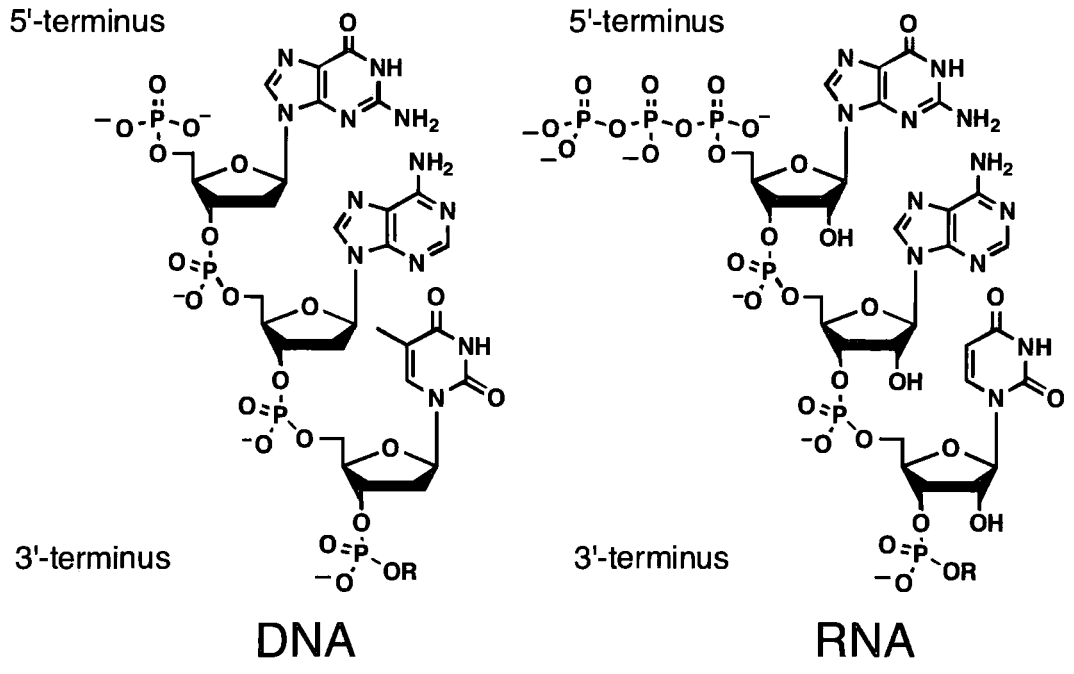


Fig:1.5 Structure of DNA and RNA showing the sugar phosphate backbone linked *via* phosphodiester bonds between 5′ and 3′ positions of the nucleotide monomers. **R** = further DNA/ RNA nucleotides. Each nucleotide shown with its hetrocyclic base attached, in a DNA duplex structure the polar groups present on these bases would base pair with the opposing anti-parallel strand.

1.3 Double stranded structure

The specific base pairing capability of oligonucleotides determines the structural interaction between DNA molecules. Hydrogen–bonds from adenine to thymine and cytosine to guanine are responsible for both the secondary structure of DNA and a great deal of the molecules ability to replicate genetic information.

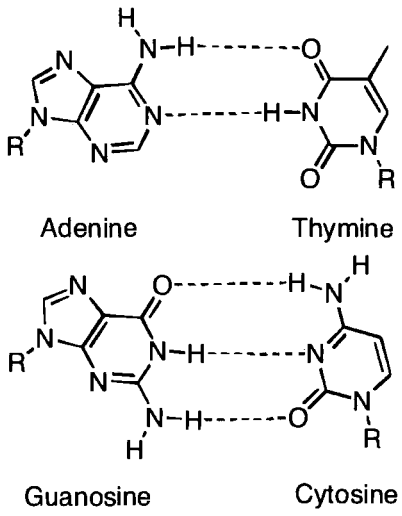


Fig:1.6 Watson-Crick base pairing between Adenine-Thymine, Guanosine-Cytosine.

In the classic Watson - Crick Model two complementary DNA strands interact in an anti-parallel fashion to form a right-handed double helix. The only links between the two strands being the hydrogen bonding between base pairs, although individually weak in total these interactions are enough to hold the two strands in a stable duplex. In RNA thymine is replaced by the similarly structured uracil, which also base pairs with adenine.

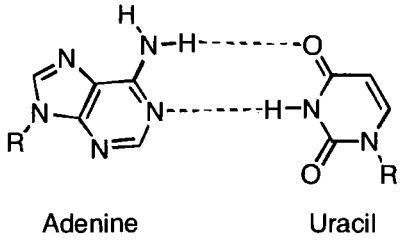


Fig:1.7 Base pairing between Adenine and Uracil RNA bases.

Unlike the double-stranded nature of DNA, RNA usually only exists as a single strand, however base pairing between complementary regions of the same strand is still possible. Often this leads to the formation of complex secondary structures such as loops, hairpin turns and hammerhead formations. Such complex structure formation is especially important for the catalytic behaviour observed in ribozymes.

1.5 Chemical synthesis of RNA

The highly efficient (each step occurs at ~99 % conversion) solid supported synthesis of DNA oligonucleotides of any desired sequence has been possible through the use of phosphoramidite coupling methods for many years. Recently the development of automated techniques for the synthesis of both DNA and RNA sequences has resulted in the opportunity for 'made to order' oligonucleotide manufacture. There are considerable limitations on the length of oligonucleotides available (commonly a maximum of 80 nucleotides) although it is possible to acquire sequences containing modified or non–canonical nucleotides. Owing to the specialised nature of the nucleic acid derivatives the costs of custom manufacture are high, especially in comparison to enzymatic methods of RNA synthesis (Section 1.5). This can preclude a synthetic route where a large–scale quantity is required, such as the production of a RNA sequence library for a SELEX experiment (~ 10^{14}–10^{15} random sequences)[1].

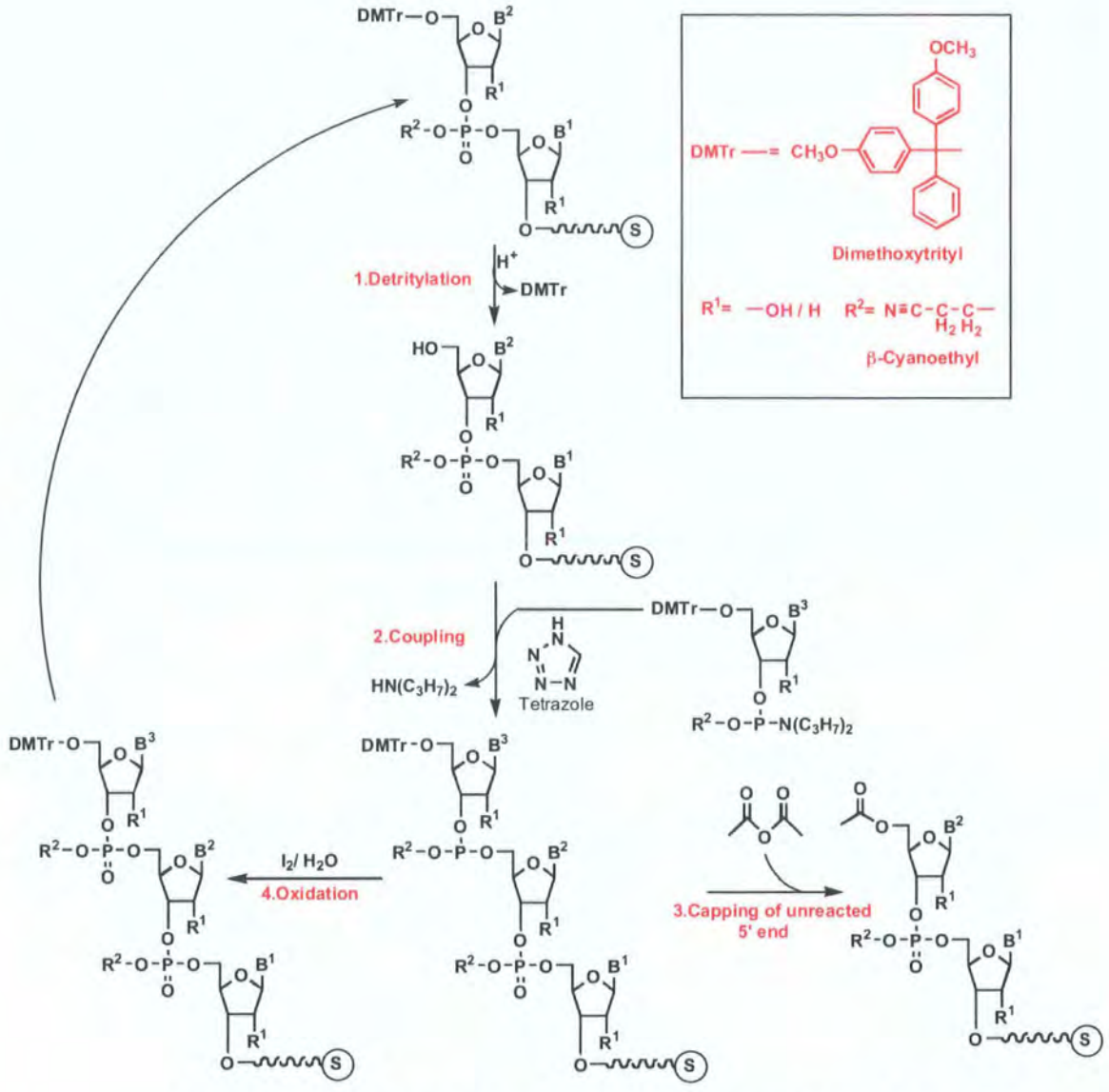


Fig:1.8 Schematic for solid-supported synthetic cycle for oligonucleotide production. Unlike biosynthetic pathways the phosphoramidate method of synthesis takes place in the 3′ to 5′ direction[2].

Even if the cost or length of oligonucleotide required is too high, the chemical synthesis of oligonucleotides by the phosphoramidite method can still be a key part of RNA production. In these cases a DNA template dependent RNA polymerase, such as T7 RNA polymerase, can carry out RNA transcription (see Section 1.5), where a sample of the DNA template is produced by chemical synthesis and amplified by the polymerase chain reaction (PCR).

1.5 Transcription of RNA sequences

Transcription is the process by which RNA molecules are biosynthesised *via* enzyme mediated polymerisation reactions. Template dependent RNA polymerase enzymes produce RNA chains of a base–sequence complementary to that of a DNA template. The chemical reaction catalysed by the enzyme is a phosphoryl transfer reaction, which forms a phosphodiester bond between nucleotides. Melting of the double helix of the DNA template to form two single strands facilitates the base pair binding of nucleotides, which are then linked via phosphodiester bonds by the polymerase enzyme. This melted region is known as a transcription bubble and is stabilized by the polymerase enzyme (Section 1.11)[3-6].

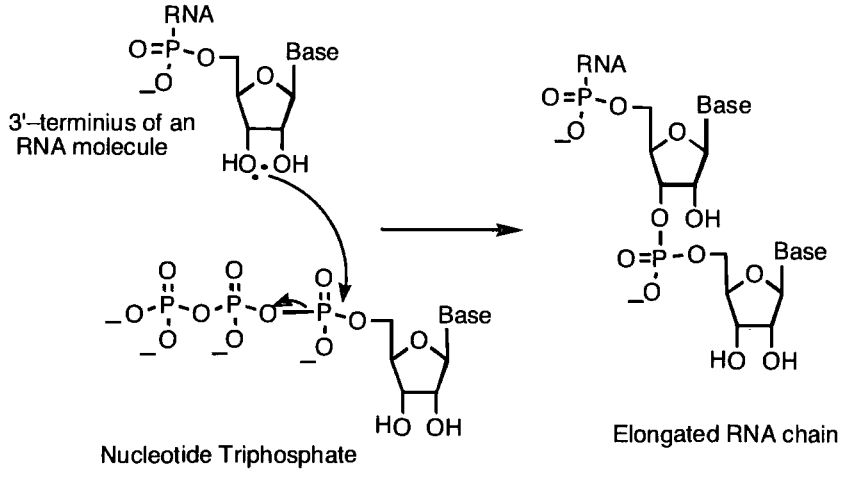


Fig: 1.9 Detail of the phosphoryl transfer reaction catalysed by polymerase enzymes. The reaction is postulated to begin with nucleophilic attack by 3′–OH of the RNA chain at the α-phosphorus of the nucleotide triphosphate. Subsequent elimination of pyrophosphate results in formation of phosphodiester bond and elongation of the RNA chain by one nucleotide.

Of the RNA polymerase enzymes available perhaps the most common to be used with novel nucleotide initiators is T7 RNA polymerase (Section 1.26). As the structure and function of this T7 RNAP was crucial to the success of our research I felt it prudent to dedicate a number of sections to detailing the structure and proposed mechanism of action of the enzyme.

1.6 Features of T7 RNA Polymerase mediated transcription

T7 RNA Polymerase (T7 RNAP) is a ~ 107 kD, template dependent, RNA polymerase of bacteriophage origin[5]. The primary function of the enzyme is to catalyse the phosphodiester bond formation between nucleotides during RNA transcription[7]. RNA transcripts are produced in the 5′→3′ direction, complementary to a template oligonucleotide sequence that is usually a dsDNA chain, although transcription with ssDNA and RNA templates has also been observed [8]. The RNA yield of the transcription reaction is at its highest when the DNA template contains a T7 promoter sequence within the template strand, this duplex sequence of nucleic acids creates a recognition site for the polymerase to bind to and dictates the base at which transcription will initiate[3-6]. Using T7 mediated transcription reactions RNA sequences thousands of nucleotides in length can be produced from both biosynthesised and chemically synthesised DNA templates[9-14].

The transcription reaction is imperfect in that the RNA transcription often 'overruns' and produces transcripts with an additional base at the 3′–terminus, relative to the sequence coded for by the DNA template. These transcripts are referred to as being n+1 in length, with those of the exact length desired for are of n length[3,6]. Many of the products of the transcription reaction are 'abortive transcripts' 2–6 base pairs in length that occur as a consequence of the enzyme's mechanism of action.

1.7 Features of the T7 promoter sequence

The T7 promoter region of a DNA template is a specific sequence of nucleic acids which directs RNA transcription *via* T7 RNA polymerase enzyme[15]. The promoter sequence shown in Fig:1.10 is that of the first T7 promoter to be identified, it both enhances the rate and productivity of RNA and directs transcription to be initiated by a guanosine species. Hence, when using this promoter within a DNA template the resulting transcripts almost uniformly posses a guanosine nucleotide at their 5′–terminus. More recently the T7 ϕ2.5 promoter sequence, which increases the productivity of transcription while directing initiation of RNA transcripts with adenosine species, has been used by Huang *et.al*[16]. To

gain an overview of the enzyme's properties and function, however, we will concentrate on the use of the more fully characterised G directing promoter.

Fig:1.10 Promoter sequence of T7 RNA polymerase. Bases -17 to -6 are important for binding the promoter sequence to the polymerase enzyme. The **-4 to -1** section helps facilitate the 'melting' of the template duplex by the enzyme.

The T7 promoter can be categorised as a tripartite structure, each section of which is important for the binding and use of the template sequence. The duplex sequence from –17 to –6 is critical for specific binding to the polymerase as revealed by numerous interactions with bases in this section of the RNAP-promoter complex crystal structure (Fig:1.10)[4,5]. The nucleotides in positions –4 to –1 of the promoter are hypothesised to be integral to the melting of the template-promoter duplex. The third section of the promoter sequence lies directly downstream from the transcription start-site (base-pairs +1 to +6), these are the first bases of each RNA molecule to be transcribed. Mutation of the base pair at positions +1 (with a reduced effect when mutation occurs at position +2) results in a reduced rate and yield of transcription and increase in the production of short (abortive initiation)[6]. The significant effect this single base mutation has on the products of transcription reactions confirms the overall guanine directing behaviour of this T7 promoter sequence.

1.8 Recognition of the promoter sequence

The ability of T7 RNAP to recognize the specific T7 promoter sequence and initiate transcription is derived from the structural characteristics of the polymerase[17]. The Palm, Thumb, Fingers and N-terminal sub-domains, named as a structural analogy to a right hand, are four regions of the enzyme that undertake key roles in the catalytic cycle and selectivity of the polymerase reaction. Sequence-specific recognition of the duplex promoter by T7 RNAP is accomplished by both direct and indirect base-recognition.

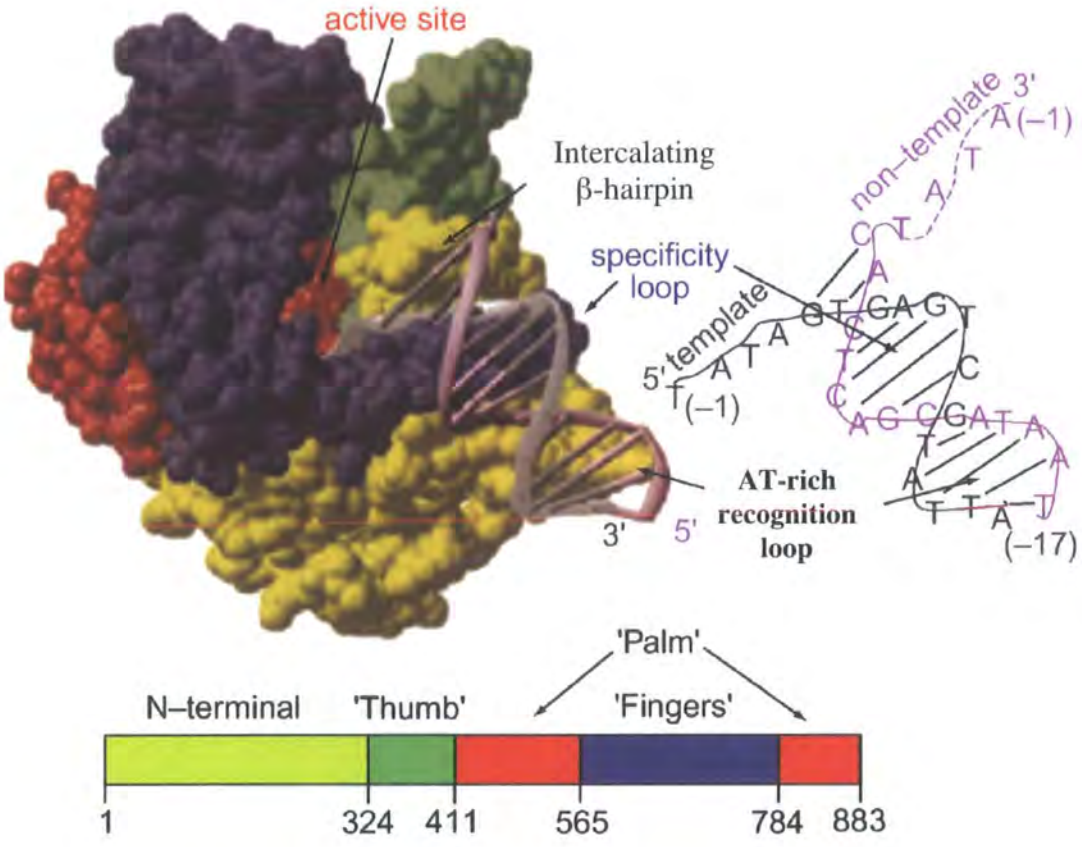


Fig:1.11 Crystal structure of T7 RNA polymerase showing colour-coded structural areas and interactions with T7 promoter region of the DNA template [17].

1.9 Direct template recognition

The T7 RNA polymerase enzyme possesses a number of structural features that are thought to be involved in the recognition of the precise nucleotide sequence that of the promoter section of the DNA template. One such structure is an anti-parallel β-ribbon, or specificity loop (residues 739-770), which is a sub-structure of the 'Finger' sub-domain (purple section in Fig: 1.11) which inserts into the major groove of the DNA template, achieving direct recognition of the specific base sequence [5]. The β-ribbon presents four amino acid residues for interaction with exposed base pairs in the promoter sequence, the crystallographic diagram showing the interactions can be seen in Fig:1.11,

a.a residue	Base
Arginine (756)	Guanosine (-9)
Glutamine (758)	Adenine (-8)
Arginine (746)	Guanosine (-7)
Asparagine (748)	Guanosine (-11)

Table:1.1 Amino acid residues of the β-ribbon sub-structure of T7 RNAP paired with the bases of the template strand they interact with. Interactions are all hydrogen bonds, usually between the keto or imino groups of guanine/adenine groups and the various amino acid residues. An interaction between Asparagine (748) and Guanosine (-11) is worthy of special note as the hydrogen bond occurs through a water molecule as shown in Fig:1.12.

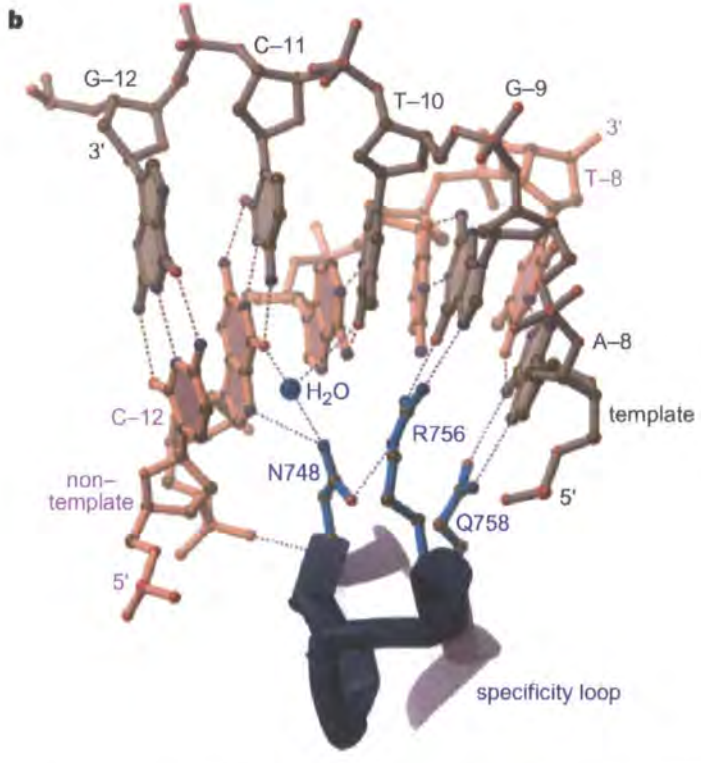


Fig:1.12 Crystal structure showing detailed interactions between the specificity loop sub-structure of T7 RNA Polymerase enzyme and both sense and anti-sense strands of the T7 promoter template [4].

1.10 Indirect template recognition

It is hypothesised that indirect (*i.e.* non-bonding) recognition of the DNA promoter sequence is achieved by the insertion of a flexible surface loop (residues 93-101) from the

N-terminal sub-domain in to the template's minor groove[4]. This formation recognises the A-T rich region in the minor groove at base pairs –17 to –13. A subsequent distortion of the phosphodiester backbone forms a wider and shallower minor groove compared to a B-form DNA helix. It is possible that this distortion, made possible by the inherent flexibility of the A + T-rich sequence, is another mode of recognition by T7 RNAP[4].

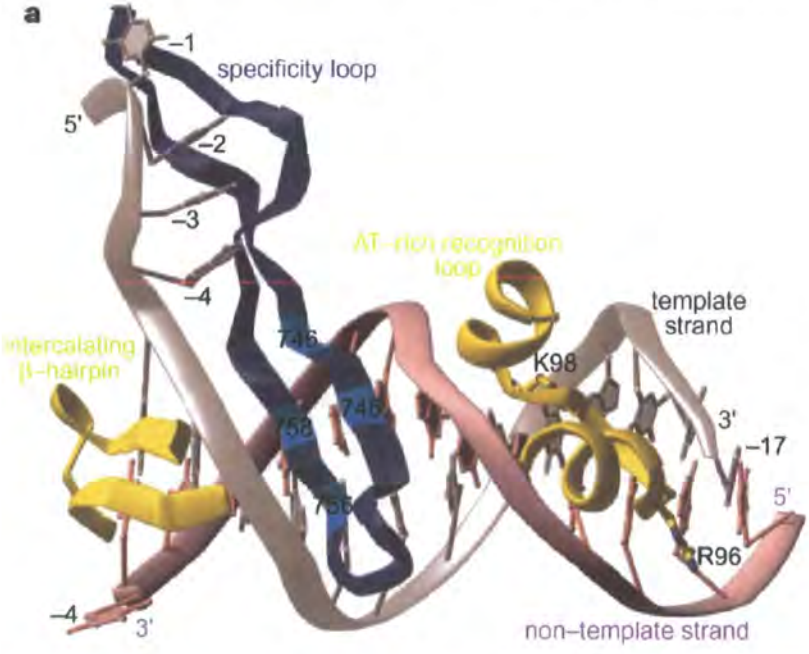


Fig:1.13 Crystal structure showing indirect recognition of the promoter region of the DNA template by N-terminus and Finger domains. [4].

1.11 Transcription bubble formation

The specific loop and N-terminal sub–structures are also important in the formation of an upstream transcription bubble[4]. A transcription bubble is created when the two complementary strands of a DNA template duplex unwind over a short number of base pairs to expose the single strands to the active site of the polymerase enzyme and allow complementary binding of free ribonucleotides. The N-terminal domain and specificity loop facilitate the melting of the promoter duplex by providing extensive binding to the template strand, which stabilize the formation of the transcription bubble.

1.12 Metal chelation

The active site of the enzyme lies in the 'Palm' sub-domain. The critical function of the active site structure is to present two metal ions (Mg^{2+}) in the appropriate geometrical arrangement for catalysis of the phosphoryl transfer reaction[7]. The mechanism of catalysis is thought to be the same as that which takes place in DNA polymerase enzymes, although the amino–acid side–chain residues in each case are different[7].

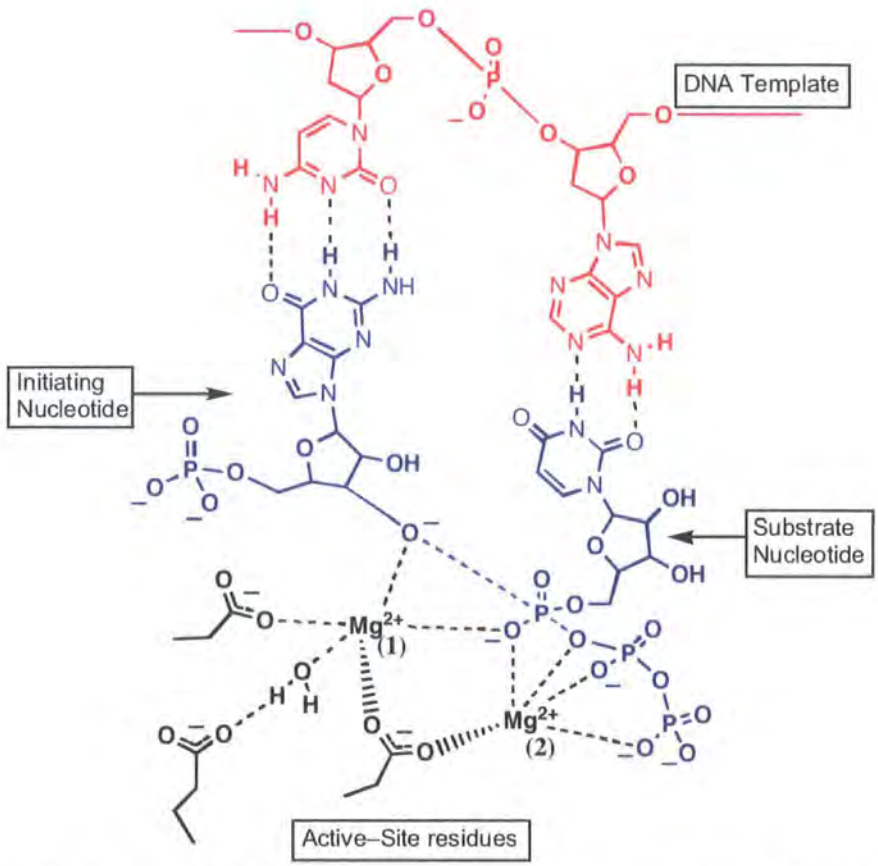


Fig:1.14 Hypothetical polymerase enzyme active site , showing amino acid residues thought to be critical in fixing the geometry of the metal ions. The carboxylate ligands are in generic positions and not intended to represent any specific. polymerase. The role proposed for Mg (1) is to lower the pK_a of the 3′-OH in order to form the oxygen anion. Mg (1) also acts to stabilize the 90° 3′-O-P-O bond angle between the apical and equatorial oxygen atoms in the transition state. Likewise Mg (2) has a part in stabilizing the pentacovalent transition state geometry and to facilitate the leaving of the pyrophosphate group[7].

1.13 Specificity of RNA transcription

RNA polymerase enzymes exclusively catalyse the formation of RNA transcripts, even in the presence of deoxynucleotides. Mutation studies of the wild type T7 RNA polymerase enzyme have proposed that the selectivity of the transcription reaction for RNA results from hydrogen bonding interactions between amino acid residues in the active site and ribonucleotides. A tyrosine residue interacts with the 2′–OH of RNA nucleotides, binding them in the active site. As deoxynucleotides do not posses a 2′–OH, no similar interactions with the tyrosine residue are possible, resulting in the binding affinity to the active site being weaker than that observed in RNA. The variation in binding affinity is observed as a preference for RNA nucleotides being used in transcription. [18].

1.14 Preference for initiation with guanosine

As well as its role in catalysis the 'Palm' domain of T7 RNA polymerase was also hypothesised to be responsible for the enzyme's selectivity towards polymerisation initiators[15,19]. Activity studies of T7 RNA polymerase reactions wherein the primer nucleotide was varied were initially thought to have demonstrated the enzyme's distinct preference for initiating polymerisation with a compound that incorporates a guanine base. Crystallographic data has been used to form the hypothesis that hydrogen bonding between a histidine residue in the 'Palm' of T7 RNAP (H784) and guanine's 2–amine group is responsible for the observed affinity for guanine in the active site[15,19]. However, the assignment can be contended since the development of the adenosine directing T7 ϕ2.5 promoter sequence, which indicates that the enzyme's observed preference is a product of the promoter sequence used, and not a property of the enzyme's structure.

1.15 Abortive transcripts

The production of short length transcripts 2 – 6 nucleotides in length is an unwanted side reaction of T7 RNAP mediated RNA synthesis. These oligonucleotides are termed abortive transcripts as they result from initiated transcriptions reactions that fully detach from the active site before completing the elongation stage of polymerisation, terminating reaction prior to transcription of the full sequence[3,6]. A common position for abortion to

occur is at approximately base +6 (as encoded by the DNA template), it is thought that the steric strain within the active site reaches a maximal limit at this point, resulting in the abortive detachment of the transcript, or partial detachment and continued transcription.

1.16 Variation in the 5′–group of initiator nucleotides

The standard initiator for RNA transcription is guanosine triphosphate, also used during the elongation phase of polymerisation. The use of guanosine and guanosine monophosphate as primers in transcription reactions produces RNA yields comparable to those observed for guanosine triphosphate. Research has shown both guanosine and GMP have a higher affinity for the T7 RNAP active -site than GTP.[20]

Primer	K_m (nM)
GTP	0.6
GMP	0.33
Guanosine	0.46

Table:1.2 T7 RNAP Michaelis constant values for guanosine primers[20]

The increase the in affinity for the enzyme active–site observed as the size of the guanosine derivatives is reduced suggests that the large structures removed undergo unfavourable interactions with the active–site. A lack of interaction between the enzyme's active site and the 5′–group of initiating nucleotides gives a wide scope for modification of initiators in this region, as will be discussed later (Section 1.26).

1.17 Bioconjugation

Bioconjugation reactions involve the attachment of a species of natural or synthetic origin to a biomolecule *via* chemical methods. Commonly the linking of the two compounds creates a novel complex that still retains the properties of the individual components, *i.e.* in linking a fluorescent group to an enzyme molecule it would be undesirable to inhibit

the enzymes activity, or disrupt the chromophore of the reporter group. Ideally conjugation to biomolecules should be:

- chemo/regiospecific,
- high yielding,
- stable to chemical/environmental conditions.

By meeting all these criteria the bioconjugation can produce a homogeneous population of derivatised biomolecules. Hence processes that rely on measurements and/or selections enabled by the presence of a conjugated species can be assumed to be representative of the entire biomolecule population.

In general biopolymer molecules, such as nucleic acids and peptides, can present difficulties when attempting to attach species with high chemo/regiospecificity as in general biomolecules lack chemical groups possessing distinct reactivities. To circumvent this deficiency, technology has been developed to introduce novel reactive groups to molecules such as RNA and DNA without interfering with the bioconjugated molecule's functionality.

1.18 Why bioconjugate to RNA?

The site-specific modification of RNA chains has become desirable in a number of areas of research, *e.g.*:

- The predicable and specific binding nature of nucleotide bases has created interest in the use of complementary oligonucleotide probes for detecting genetic markers for disease.
- The discovery of catalytic activity in some RNA chains (ribozymes) has created an interest in the evolution and analysis of such molecules (for ribozyme selection studies it is necessary that the RNA chain of interest be distinguishable from 10^{14}–10^{15} of other oligonucleotides present)[1,9,11,21].

The types of species that have been attached range from fluorophore compounds to provide a non-radioactive method of RNA detection, reagent molecules in studies of RNA

catalysis such as the anthracene derivative shown in Fig:1.15, and affinity tags that allow the separation of bioconjugated and unmodified species[1,22-24]. The 5′–terminus of RNA chains is a popular site for modification owing to its unique nature and position in the molecule:

1. Being positioned at the end of the RNA sequence means that even in sequences proven to have complex secondary structures, groups at the 5′–position have proven to be readily available for bioconjugation[1,10-12,21,25-30].
2. As a consequence of transcription occurring in the 5′ to 3′ direction in T7 RNAP, the highly regiospecific incorporation of novel groups at the 5′ position has been demonstrated for a number of chemically diverse groups (Section 1.26)[10,26-28,31].
3. The 5′ to 3′ direction of transcription and the G directing nature of the T7 promoter sequence has the potential to produce in RNA transcripts possessing a homogeneous base at the 5′–terminus. Run–off transcription where an additional base is added to the 3′–terminus of transcripts (n+1 length) results in the formation of a heterogeneous terminus.

Modification of RNA has found broad applications in studying RNA structures, mapping RNA-protein interactions, and *in vitro* selection of ribozymes.

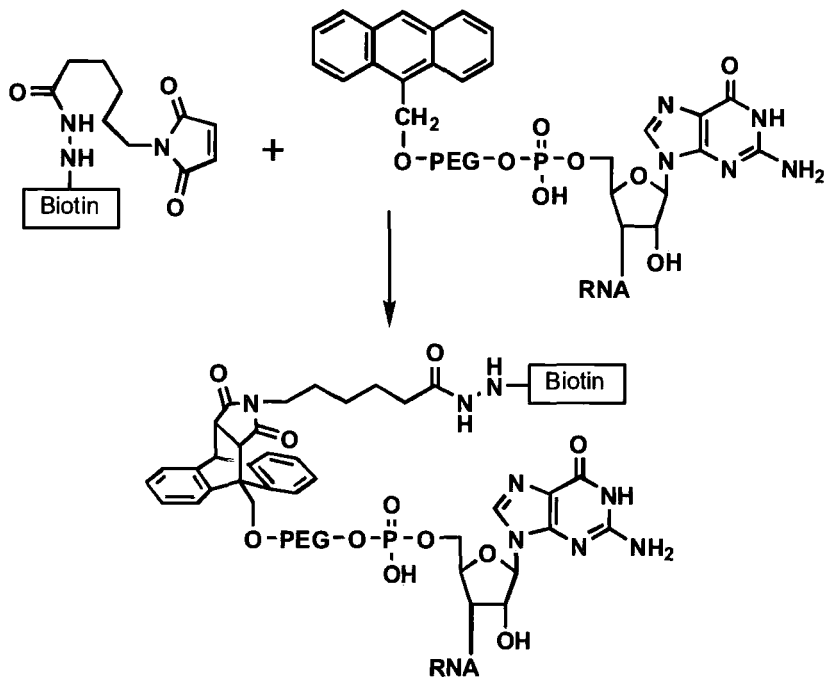


Fig:1.15 The bioconjuation of anthracene to the 5′–terminus of RNA allowed Jäschke and Seelig to select and isolate an RNA sequence Diels–alderase ribozyme) capable of catalyzing a Diels Alder reaction between the tethered substrate and a maleimide derivative[10].

1.19 Chemical modification of RNA

A commonly used method of introducing reactive groups to RNA chains for use as 'handles' for bioconjugation is to undertake posttranscriptional modification of the nucleic acids subunits. As these methods utilize familiar synthetic methods I have gathered them under the heading of 'chemical modification'. The chemical modification of nucleic acids at specific sites within long chains allows the introduction of various groups, such as amines or thiols, into polynucleotide molecules. Once a functional group has been introduced it can be used as a 'handle' for highly regiospecific bioconjugation reactions. The chemical modifications are not limited to enzymatically produced RNA or DNA strands and can also be applied to chemically synthesised nucleic acids [32]. The first step in modification is usually the attachment of a spacer component to a marginally reactive group in the RNA molecule, the reactivity this species introduces can then be used to couple another, more useful, molecule to the RNA. The following sections will

describe a number of positions within RNA molecules that have been utilized for the attachment of functional groups.

1.20 Base modification

One of the most common sites used in nucleotide modification are the DNA/RNA bases themselves. Numerous methods have been developed for the introduction of a wide range of compounds to the various bases *via* both nucleophilic and electrophilic substitutions. As the structures of pyrimidine and purine bases differ significantly separate methodologies have been developed for the attachment of functional groups to each substructure. In the case of both pyrimidine and purine bases a number of sites exist that are susceptible to nucleophilic attack, while other locations can be targeted by electrophiles[32].

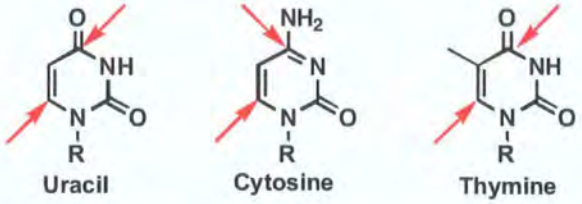


Fig:1.16 The three pyrimidine bases showing common sites of nucleophilic attack[32].

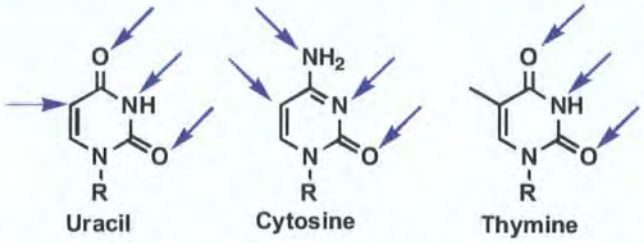


Fig:1.17 The three pyrimidine bases showing common sites of electrophilic attack[32].

Although many potentially reactive sites on pyrimidine bases exist, the intrinsic reactivity of pyrimidines themselves is low, resulting in low specificity and therefore inefficient bioconjugation. In general the rates of reaction are low to match the poor yields, especially in aqueous environments[33]. The purine bases (adenosine and guanine) also have several sites available for attack by nucleophilic and electrophilic reagents, though their reactivity towards nucleophiles is even poorer than their pyrimidine counterparts.

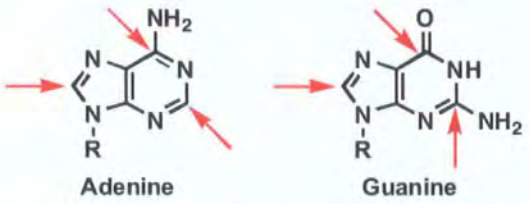


Fig:1.18 The two purine bases showing common sites of nucleophilic attack[32].

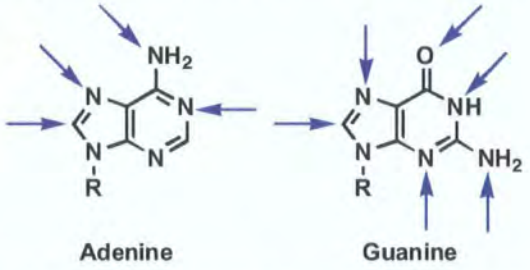


Fig:1.19 The two purine bases showing common sites of electrophilic attack[32].

Modification of RNA bases, whether chemical or enzymatic, is commonly carried out after the synthesis of the oligonucleotide target has been completed. By following this method the nature of the modification that is to be made can be controlled, very little control is achieved over the positioning of the modification *i.e.* it could be at a terminus, or any position in the body of the sequence. Therefore bioconjugation to the base sub–structures by chemical methods cannot be employed where conservation of the functionality or structure of the oligonucleotide is a requirement of the experiment. A commercially available base modification system is the ULYSIS labelling kit, which allows the attachment of a stable fluorescent label to oligonucleotides *via* guanosine residues[34].

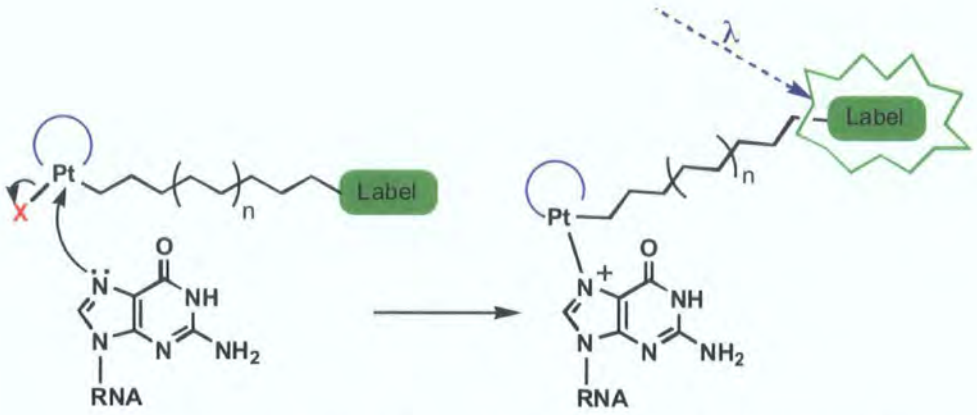


Fig:1.20 Diagram showing the general mode of action of the ULYSIS nucleotide labelling system. A fluorescent label is often used, although, a variety of reporter species could be employed[35].

1.21 Ribose modification

In addition to modification of the bases, bioconjugation reactions can also be carried out using the ribose substructure of nucleotides. As with the various bases there are a number of sites present on the ribose that could potentially be open towards modification, these sites are essentially limited to 5′,2′ and 3′ hydroxyl groups, of which the 5′ is most reactive (*i.e.* nucleophilic) though not by a large margin. With the three hydroxyls displaying almost equivalent nucleophilicities the selectivity of any chemical reaction utilizing these groups becomes a major factor when choosing an attachment site. In order to circumvent these issues chemical modification of the sugar group and any subsequent bioconjugation reactions are commonly carried out post-transcription when only the 2′,3′-hydroxyls are available (owing to the presence of a triphosphate group at the 5′-terminus). Modification of the terminal groups of oligonucleotides has the distinct advantage over base–modification as there is less chance an attached moiety will disrupt any base pairing interactions essential to the activity of the oligonucleotide, *e.g.* secondary structure in ribozymes.

1.22 Chemical modification of the 3′–terminus of RNA

One of the most successful methods for modification of the 3′-terminus of RNA transcripts is through cleavage of the 2′,3′ C–C bond to form a bioconjugation 'handle' in the form of two aldehyde groups (see Fig:1.21)[36].

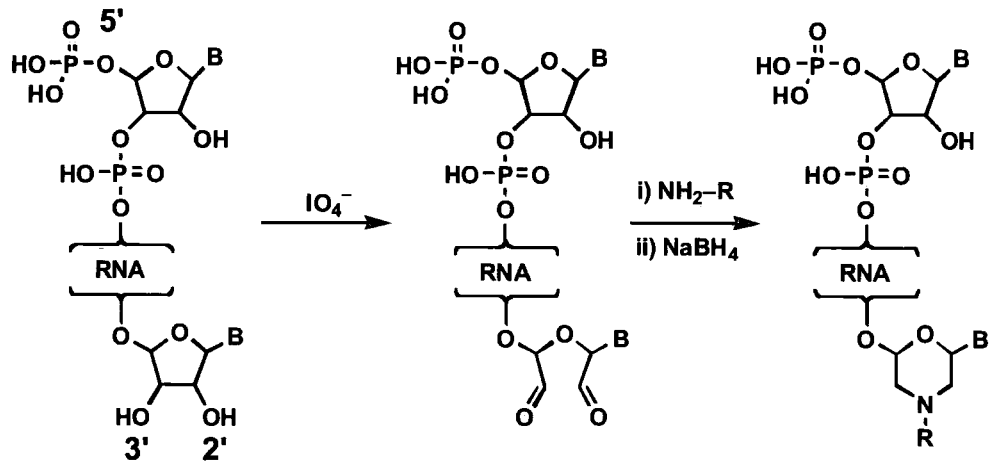


Fig:1.21 Modification and subsequent bioconjugation to the 3′-terminus of RNA, where R can be a wide range of reporter molecules. The use of an amine nucleophile is merely illustrative; a variety of reagents can be used resulting in a range of linkages.

The formation of the 2′,3′ aldehyde groups *via* oxidation with periodate opens a route for bioconjugation to an amine nucleophile to form a stable morpholine linkage between the RNA and a desired functional group. Bioconjugating to RNA in this manner has proven both highly specific and reliable, though there is the possibility of heterogeneity at the 3′–terminus. A downside to modification of the 3′–terminus through the cleavage methodology is the need to carry out multiple reactions on small amounts of material that could be lost during work-up stages.

1.23 Enzymatic 3′–terminus modification

In addition to post–synthesis chemical modification, the 3′–terminus of oligonucleotides can also be modified through the enzymatic attachment of pre–modified nucleic acids. This technology was first developed to transfer an α–^{32}P labelled nucleotide to the

3′–terminus of an oligonucleotide fragment, but has been expanded to include the attachment of affinity tags, such as biotin, and reactive 'handles' for bioconjugation. The enzymes used to make these modifications vary with each application, though usually include a polymerase and a transferase.

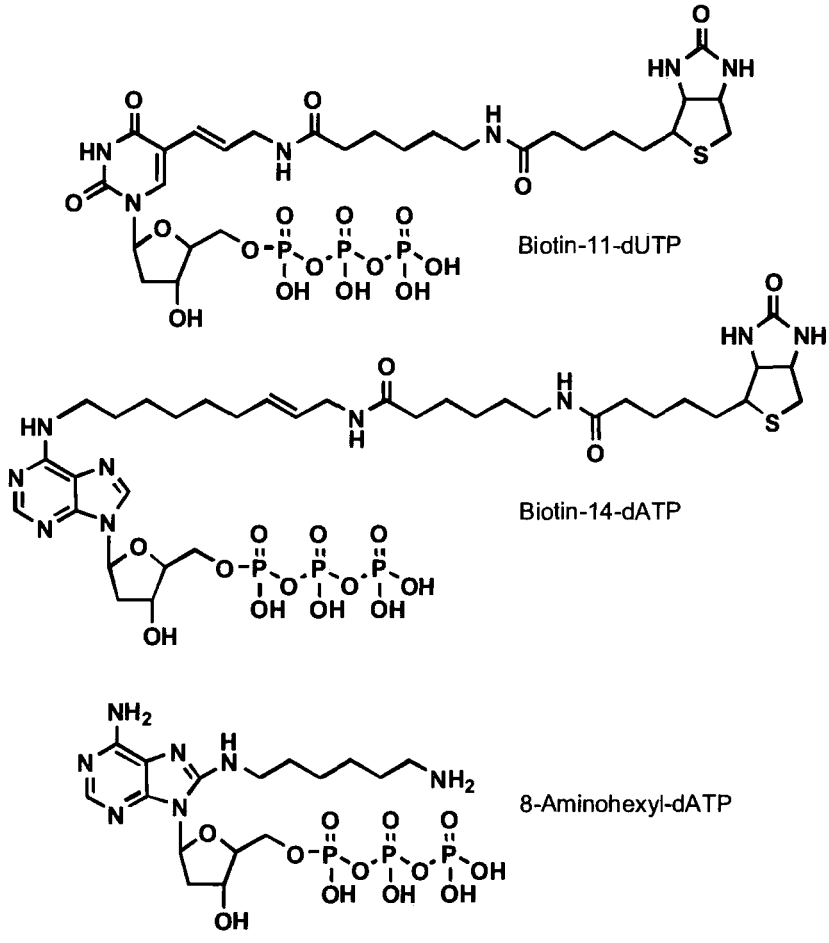


Fig:1.22 Three examples of modified deoxynucleic acids that can be enzymatically transferred to the 3′–terminus of oligonucleotides[37,38].

1.24 Chemical modification of the 5′–terminus of RNA

In addition to the post transcription modification of the 3′–terminus the 5′–terminus of RNA is also available as a site for bioconjugation, a commonly used method for achieving this transformation is shown in Fig:1.23. Using this strategy EDC and imidazole are used to form a reactive phosphorimidazolide with the phosphate group present at the 5′–terminus of the oligonucleotide. This electrophilic species is then

subjected to nucleophilic attack by ethylene diamine to form a terminal amine which can then be used as a bioconjugation 'handle' [39-41].

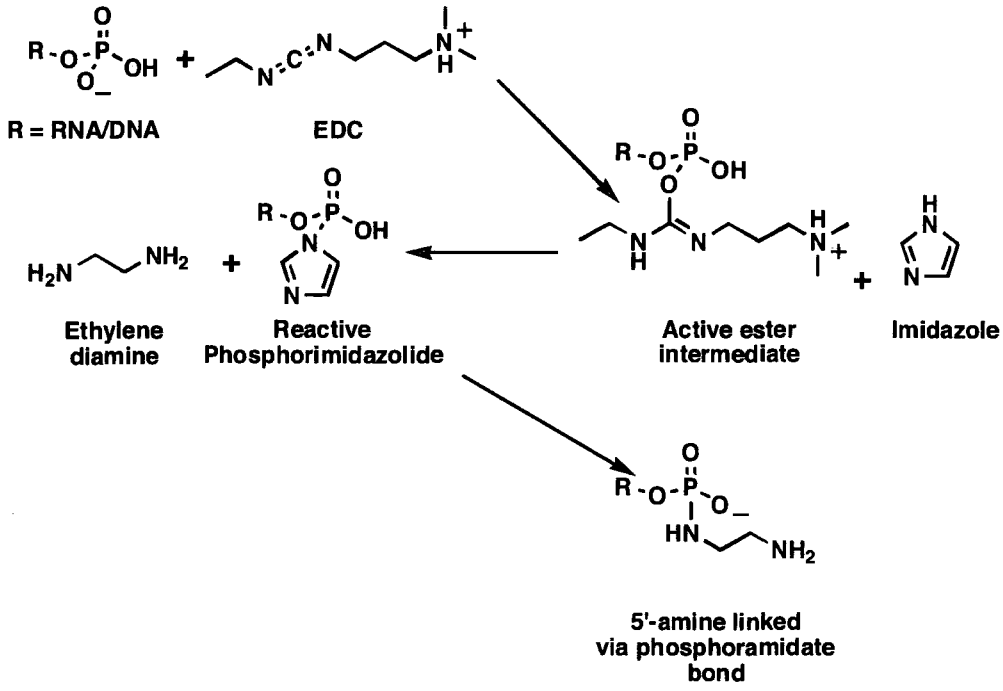


Fig:1.23 Oligonucleotides containing a 5′-phosphate group can be reacted with EDC in the presence of imidazole to form an active phosphorimidazolide intermediate. This derivative is highly reactive with amine nucleophiles, forming a phosphoramidate linkage[39-41].

The terminal amine is a reactive hard nucleophile that can be utilized as a 'handle' for further bioconjuation. However, the downside to this methodology is that the bioconjugative linkage is only as stable as its weakest constituent, the hydrolytically unstable phosphoramidate[41]. A similar methodology was reported by Chu *et.al* in which cystamine rather than ethylene diamine was bioconjugated to the terminal phosphate, the introduced terminal amine group was in turn used to attach a biotin affinity tag[43]. The modified RNA sequences were then separated from the unmodified species through conjugation to avidin. The RNA could then be retrieved through cleavage of the disulphide bond of the cystamine sub-unit by dithiothreitol (DTT).

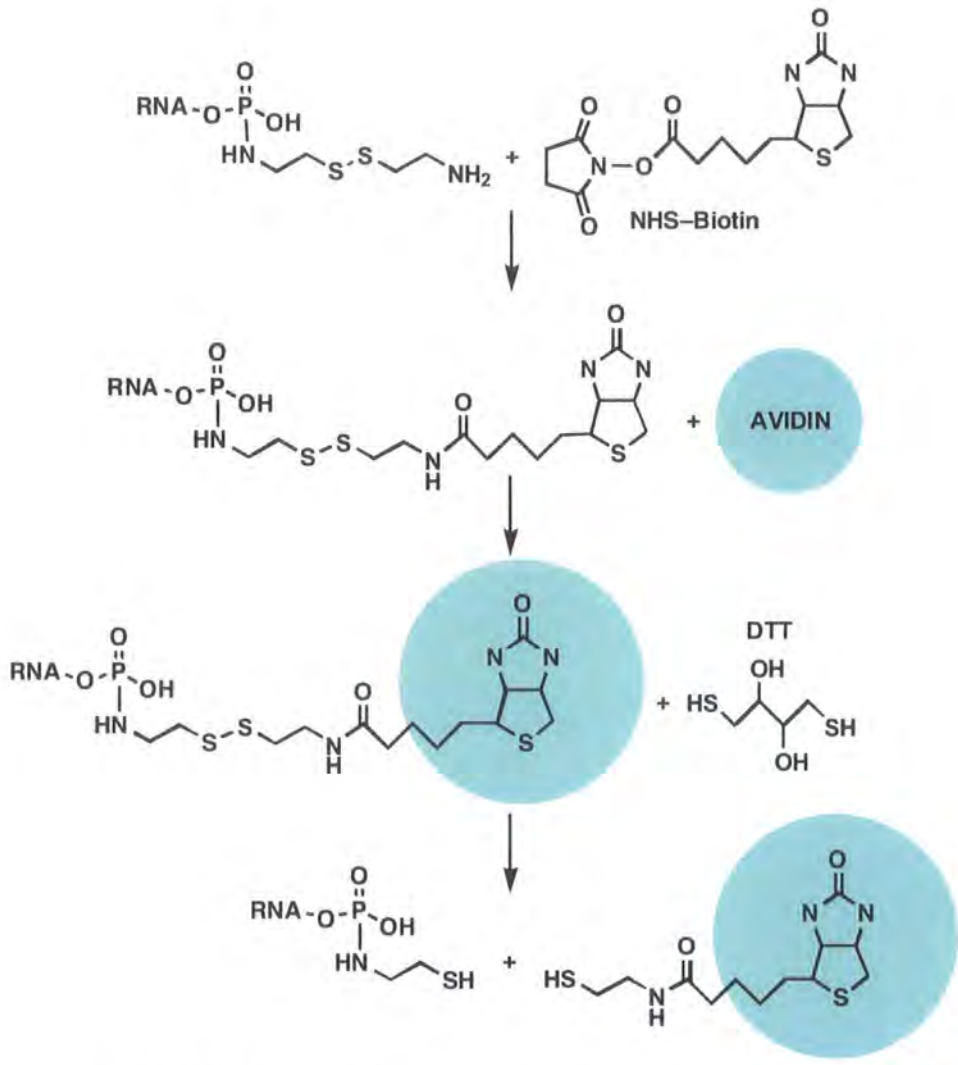


Fig:1.24 Biotinylation of cystamine primed RNA molecule with subsequent conjugation to avidin, and disulfide cleavage[32].

1.25 Introduction of 5′ sulfhydryl groups

By utilizing a reaction scheme very similar to that shown above (Fig:1.19) attachment of a reactive sulfhydryl group to the 5′–terminus of RNA can also be achieved. The presence of a soft nucleophile on the 5′-terminus of RNA opens avenues to reactions which the hard amine nucleophile would be poorly suited *e.g.* nucleophilic displacement of halide ions from alkyl halides, disulfide bond formation [40,42].

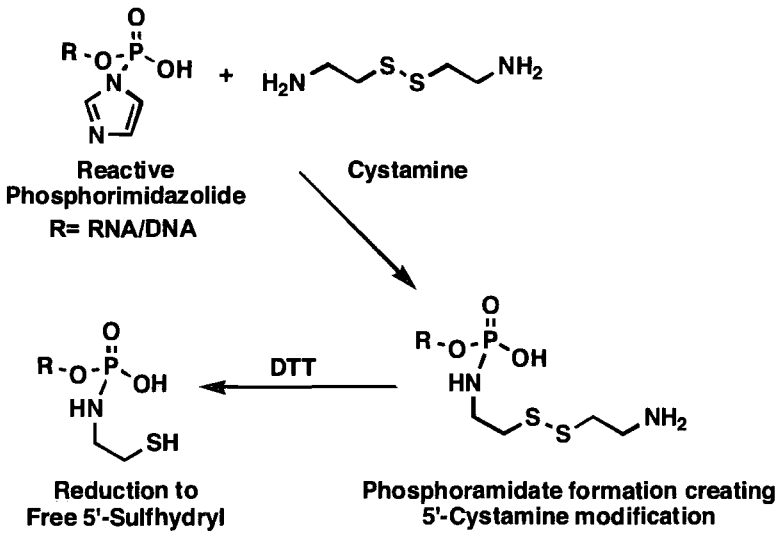


Fig:1.25 The 5'-phosphate group of the oligonucleotide may be coupled with cystamine using the EDC/imidazole reaction. This results in the formation of an amine terminal spacer containing an internal disulfide bond. Reduction of the disulfide provides a route to a free thiol, available for further derivatization[32,40,42].

The introduction of a reactive thiol group at the 5'–terminus of RNA has been successfully used by Chu *et.al* to form disulfide linkages between either two RNA sequences, or between an RNA chain and a thiol containing peptide. Ligation methods such as these could be used to attach a DNA/RNA probe to an enzyme *via* a cysteine residue[42].

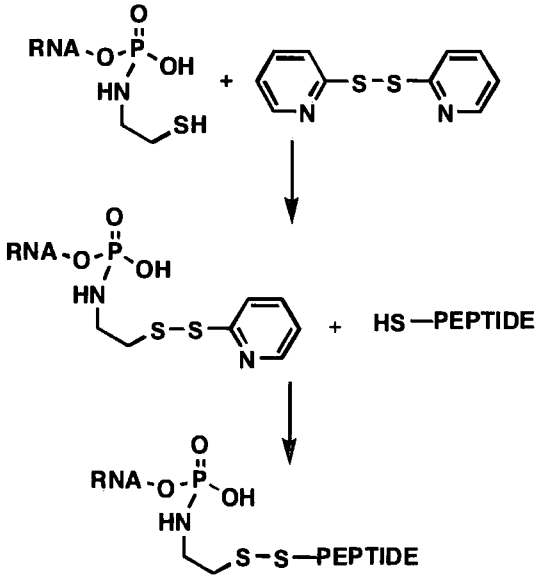


Fig:1.26 Activation of 5'–ThioRNA, followed by attachment *via* a disulfide bond to a peptide[42].

1.25 Phosphoramidate linker instability

A major factor in the applicability of RNA that has been chemically modified by the EDC methodology shown above is the pH dependent stability of the phosphoramidate linkage that has been formed. Facile cleavage of the P-N bond in conditions of moderate to low pH renders any linkage between an oligonucleotide and bioconjugated species inherently unstable, and therefore limited in use[40,42]. The unstable nature of phosphoramidates and their mechanism of cleavage will be discussed in more detail in a later chapter (Chapter 2).

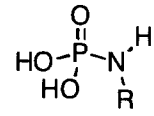


Fig:1.27 A phosphoramide group. The P-N bond is known to be unstable under moderate to low pH conditions.

1.26 Novel nucleotides and 5′–modification

A widely used and highly successful method for the modification of both long and short RNA sequences at their 5′-terminus is through the use of novel nucleotide initiators in T7 RNAP mediated RNA transcription[9-12,21,25,28,29]. Owing to the tolerance of the T7 RNAP active site towards structural modification of nucleotide substrates a variety of initiator nucleotides have been developed. The adaptability of the enzyme coupled with its ability to function *in vitro* and at low cost (the enzyme can be readily over-expressed from *E.coli* or purchased) make novel nucleotide initiator technology very appealing[43]. As noted previously when using the T7 RNA polymerase the highest yield of RNA transcripts when using the T7 promoter sequence occurs when initiation takes place with derivatives of guanosine, rather than those of other nucleobases, hence guanosine is the most common nucleotide to be modified. In recent years the development of an alternate promoter sequence that favours transcription initiation with adenosine nucleotides has led to the production of a number of novel adenosine nucleotides[16].

The sections below discuss a number of factors important to novel nucleotide design, as well as detailing some of the novel nucleotides that have been used to incorporate reactive thiol and amine groups into the terminus of RNA strands. The incorporation of a 'handle'

such as a thiol or amine make the efficient bioconjugation between RNA and a wide range of compounds possible.

1.27 Novel nucleotide design

Owing to the preference T7 RNAP shows towards initiating transcription reactions with guanosine derivatives when using the T7 promoter shown in Fig:1.8, the probability that a novel nucleotide successfully initiates transcription reactions (and therefore more RNA transcripts incorporate the novel group) is increased by creating novel nucleotide initiators from guanosine bases[19].

The tolerance of the T7 RNAP active site for novel nucleotides enables a wide range of modifications to be made, limits do exist however and the use of large groups or those that differ too extensively in nature from the usual anionic phosphate can result in a decreased total RNA yield and a decreased level of novel nucleotide incorporation[10,14,21,24,27,28]. By observing the levels of incorporation and total RNA yields resulting from transcription reactions containing guanosine derivatives modified to include a range of PEG linked functional groups Jäscke *et al* were able to make an assessment of the active sites tolerance (Fig:1.28)[44].

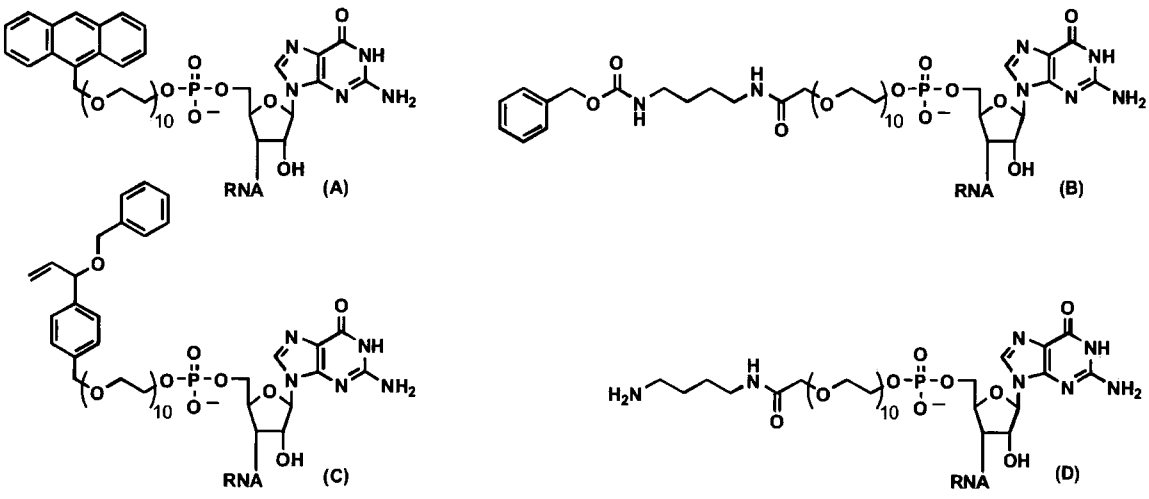


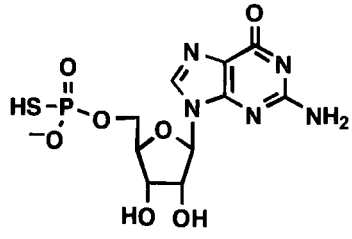
Fig:1.28 A range of novel guanosine derivatives featuring different end–groups attached to the phosphate *via* a PEG tether have been developed by the Jäschke research group. The level of incorporation observed for each nucleotide has been observed as **(A)** anthracene (90–95 %), **(B)** benzyl carbamate (68 %), **(C)** benzylally ether (90–95 %), **(D)** primary amine (49 %)[44].

All novel nucleotide initiators are used in competition with the 'natural' initiator of the transcription reaction, GTP, the presence of which is essential in the reaction mixture for the chain elongation stage of the transcription reaction. Therefore to sway the competition in favour of the novel nucleotide and increase levels of incorporation, the modified nucleotide is commonly used in excess (4 – 8 fold the concentration of GTP). Under these conditions initiation of a transcription reaction by the modified nucleotide may become more probable than with GTP. Hence, assuming no unfavourable factors make the novel nucleotide a poor initiator, levels of incorporation of the desired functional group at the 5′-terminus of RNA of over 95% can be achieved[26,28]. Care must be taken however as if used in great excess, novel nucleotides can act to inhibit RNA production, resulting in a poor total RNA yield, though potentially very high levels of incorporation. The likely reason for the observed reduction in yield is competition between the novel nucleotide (which cannot take part in any phase of polymerisation other than initiation) and GTP during the chain elongation phase of polymerisation.

1.28 Enzymatic incorporation of sulfur containing groups

The introduction of a soft nucleophile such as a sulfhydryl group to the 5′-terminus of a RNA molecule has proven to provide a very valuable 'handle' for bioconjugation[26-28,31]. The Thiol-reactive functional groups are primarily useful for carrying out thiol-disulfide exchange reactions. Formation of disulfide bond between the modified RNA and a second thiol-containing compound can be used to reversibly bioconjugate a reporter group or reagent species to the RNA transcript (Fig:1.26). The disulfide bond formed is stable under a variety of conditions, though can be efficiently cleaved using specific reagents such as DTT. This gives extra scope to the strategy of thiol introduction, as the synthetic group bioconjugated may be undesirable at a latter stage. The incorporated sulfur group can also take part in nucleophilic substitution of alkyl halides, forming linkages to synthetic groups *via* S-C bonds.

1.29 Guanosine monophosphorothioate (GMPS)



GMPS

Fig:1.29 Guanosine monophosphorothioate is a novel nucleotide commonly used to incorporate a thiol group at the 5′-terminus of RNA[27].

GMPS is a novel nucleotide initiator commonly used to introduce a reactive sulphur group at the 5′–terminus of RNA transcripts. When ionised the phosphorothioate group reveals a anionic sulphur ion, which can react rapidly with soft electrophiles. In addition to exhibiting high levels of incorporation (>80% of RNA transcribed can be primed with GMPS) the phosphorothioate has been found to have a stimulating effect on transcription reactions mediated by T7 RNAP. This observed increase in total RNA yield at low concentrations of guanosine derivatives has been attributed to an increase in the amount of guanosine species in the reaction mixture, relative to a standard (non-GMPS

containing) transcription reaction, this leads to a higher number of polymer chains being initiated and therefore more full–length transcripts are produced. RNA transcripts primed with GMPS can be modified at the 5'-end through a range of methods including disulfide bond formation (Fig:1.30)[27].

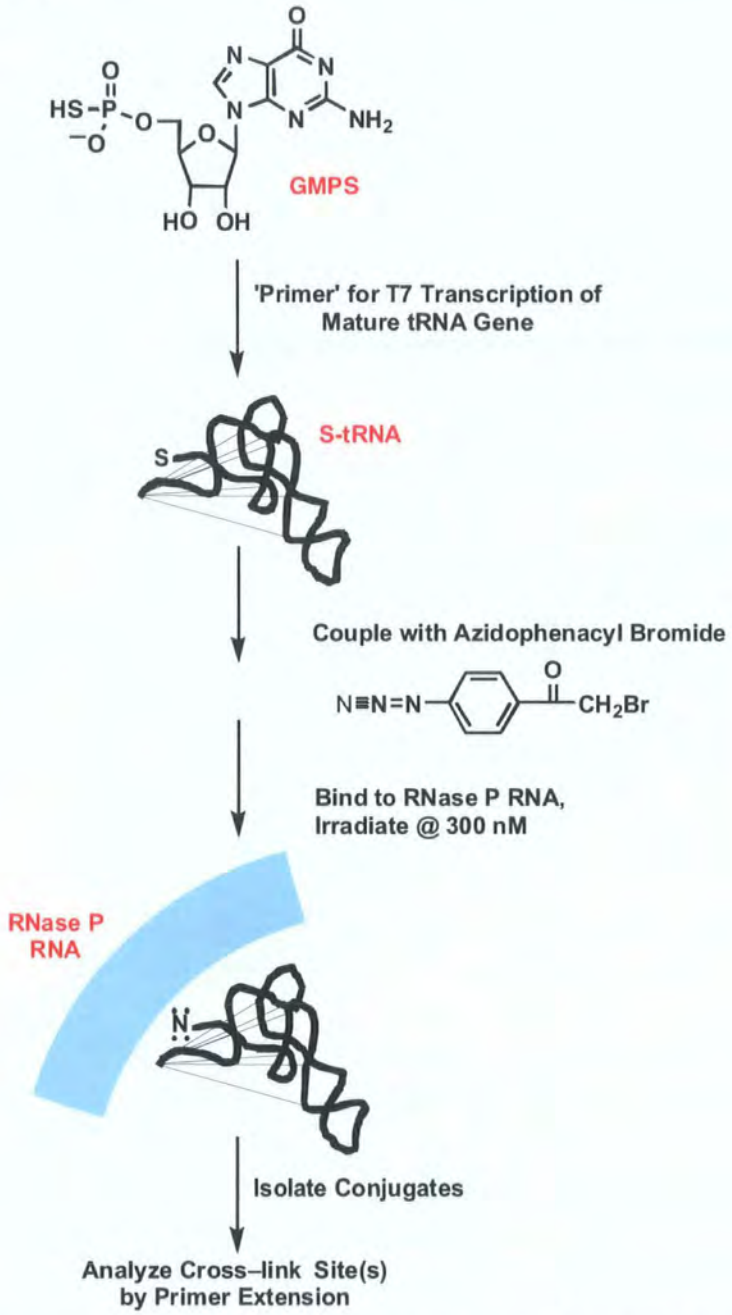
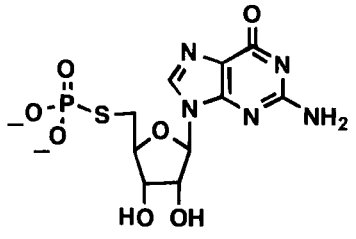


Fig:1.30 The diagram shows the method followed by Burgin and Pace were amongst the first to demonstrate the key role bioconjugation can play in the probing of enzyme structures[30].

Although common, the use of GMPS is not a perfect technology; the phosphorothioate linkage is hydrolytically unstable in conditions of low salt, temperature above 23 °C and pH >7.5, in addition to this the synthesis of GMPS is particularly unsavoury limiting the appeal of the novel nucleotide to certain user groups[26,27]. Hence, alternative routes to the incorporation of thiol groups at the 5′–terminus of RNA have been developed.

1.30 5′-Deoxy-5′-thioguanosine-5′-monophosphorothioate (GSMP)



GSMP

Fig:1.31 5′-Deoxy-5′-thioguanosine-5′-monophosphorothioate (GSMP) is used in a manner similar to GMPS[26,28].

As would be expected the inherent structural similarity between GMPS and GSMP makes the two nucleotides comparatively good initiators for T7 RNAP mediated transcription reactions[26]. Although slight, the structural difference between the two nucleotides has a significant effect on the properties of the transcripts produced, in particular the stability of the incorporated sulfur group.

Those RNA transcripts primed by GSMP possess a phosphorothioate at the 5′–terminus, as do those primed by GMPS, however the key difference occurs in the position of the sulphur atom. In GMPS the sulfur of the phosphorothioate can be ionised to form a good nucleophile as it is attached to phosphorus and a removable proton, however, this is not the case in GSMP. Transcripts primed with GSMP must therefore be activated before use by cleavage of the phosphorus to sulphur bond. Removal of the phosphate 'masking-group' reveals an ionisable thiol group.

1.31 5'-Thiol-hexaethylene glycol guanosine (Thio-PEG)

In a recent publication Jäschke and Schlatterer have detailed the synthesis and transcription efficiency of 5'–Thiol–hexaethylene glycol guanosine[31]. This novel nucleotide differs from GMPS and GSMP as it incorporates a flexible polyethylene glycol linker between the nucleic acid substructure and the novel reactive group. As shown in Fig:1.28 the inclusion of a PEG linker, though having the potential to impede transcription initiation through unfavourable steric interactions with the T7 RNAP active–site, has no effect within limits. This is sufficiently demonstrated by Thio-PEG achieving a reasonable level of incorporation into RNA transcripts (60 %)[31]. Tethering is thought to be advantageous for a number of applications *e.g.* avoiding quenching of an attached fluorophore by adjacent RNA nucleotides, or in impeding the formation of unwanted secondary structures in RNA bound directly to a matrix. It has also been claimed that a PEG linker is a beneficial addition to ribozyme selection studies, where substrate molecules are tethered to the RNA population. The benefit provided by the tether can be identified as the increased effective molarity of the bound substrate species, relative to unbound reagents. The tether length would have to be optimised to ensure substrate maintains the freedom required to enter the ribozyme active site and undergo reaction.

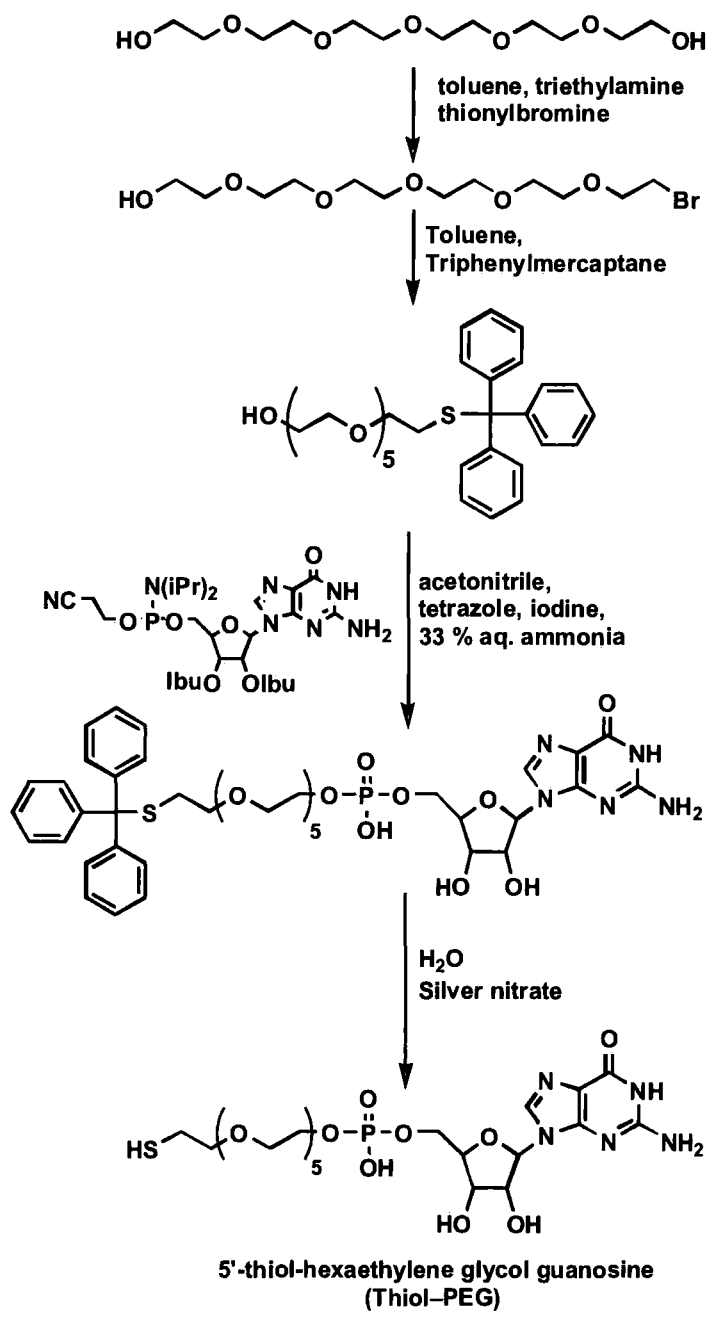
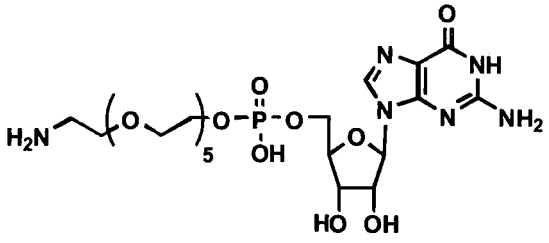


Fig:1.32 The synthetic route to Thiol–PEG[31].

The synthetic route for Thiol–PEG is quite demanding, specifically the multiple distillations of triisobutyrylguanosine–(2–cyanoethyl–*N,N*–diisopropyl)phosphoramidate from pyridine and inert atmosphere required for the coupling step. Though quite standard practice for chemical research these tasks may be too specialised and dissuade the use of Thio–PEG by a more general group of users, especially those in the biological disciplines.

1.32 Enzymatic incorporation of terminal amines

To date a number of methods exist for the incorporation of amine groups into the 5'-terminus of RNA. However, these methods suffer from either producing poor results (20 % incorporation and a reduced total RNA yield), or the synthesis of the novel nucleotide is overly complicated. Jäschke and Schlatterer's recent paper in Biochemical and Biophysical research Communications (2006) detailed the synthesis and use of 5′-Amino-hexaethylene glycol guanosine (Amino–PEG), a modified nucleotide analogous to 5′–Thiol–PEG, but possessing an amine group linked to the 5′-carbon of the ribose ring by a flexible PEG linker[31].



5'–amino–hexaethylene glycol guanosine (Amino–PEG)

Fig:1.33 The Amino–PEG novel nucleotide is a PEG linked amine derivative of guanosine; developed by Jäschke and Schlatterer[31].

This amino-nucleotide was reported to successfully incorporate into the 5′-terminus of RNA and was proven to provide a reactive amine for bioconjugation. Through using a reduced concentration of GTP in the transcription reaction, and therefore swaying the competition to initiate transcription in favour of Amino–PEG, the level of incorporation was raised to a maximum of 65%[31]. However as GTP is required for the chain elongation step of transcription, this also resulted in a reduced total RNA yield (see Table:1.3). The levels of incorporation quoted for the amino-nucleoside are not so low as to deter its use, however the synthetic route to the PEG linked amino-guanosine is quite protracted. As with the synthesis of 5′–Thiol–PEG the coupling reaction to triisobutyrylguanosine–(2–cyanoethyl–*N,N*–diisopropyl) phosphoramidate involves a high level of preparation of materials (co–distilling from pyridine) and use of an inert atmosphere; this use of specialised techniques is potentially off-putting for researchers with little or no experience in chemical synthesis.

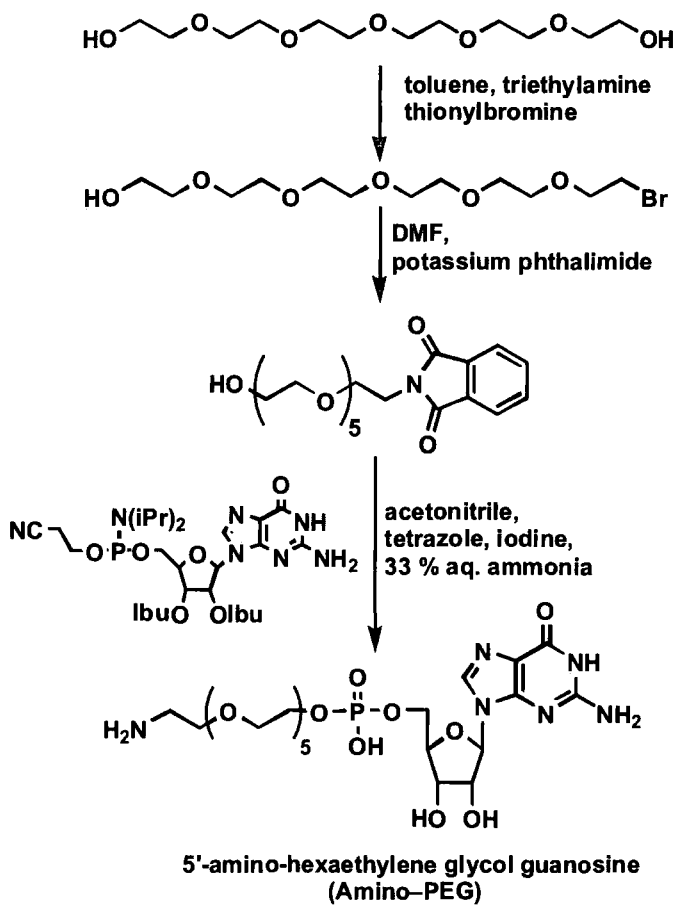


Fig:1.34 Synthetic route to Amino-PEG, 19 % yield over three steps. The synthetic steps involve the use of specialised techniques such as Silica gel chromatography and the use of an inert atmosphere for some reactions[31].

1.33 5'-amino-5'-deoxyguanosine (Amino–G)

A second method from the literature details the use of 5'-amino-5'-deoxyguanosine (Amino–G) to initiate transcription reactions[45]. The synthesis of this molecule can be achieved from guanosine in good yield (39 %) over 3 steps without need for chromatographic purification steps[46,47].

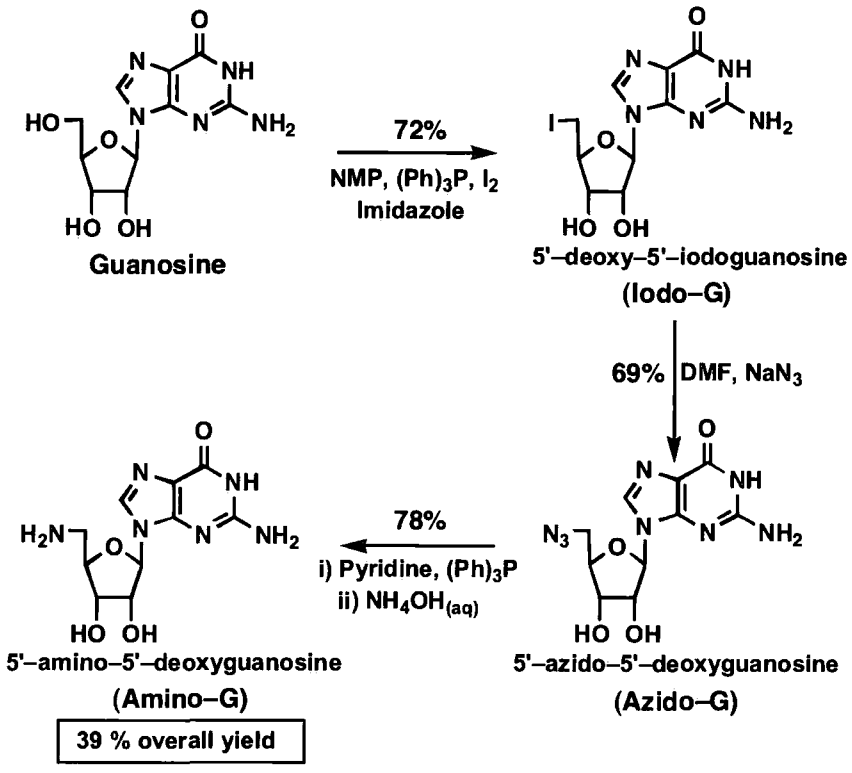


Fig:1.35 Reaction scheme for synthesis of 5'-amino-5'-deoxyguanosine (Amino–G). Yields shown are those quoted in literature. Purification of products achieved by solvent removal, filtration and washes[46,47].

During research into catalysis by the ATRib^{TL} acyl-transferase ribozyme Suga *et al* investigated the ability of Amino–G to initiate T7 RNAP mediated transcription reactions, and thereby synthesise a 5′-amino variant of the ATRib^{TL} sequence[45]. This variant was then used along with the 5′-hydroxyl terminated ribozyme to investigate the metal ion dependence of the ribozymes catalysis.

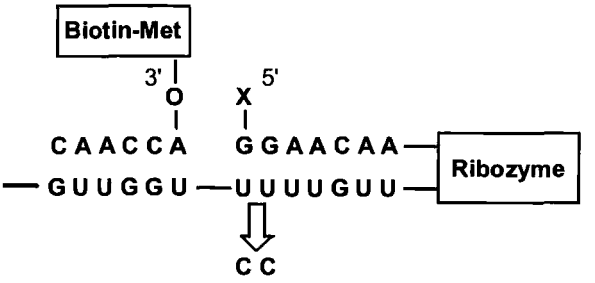


Fig:1.36 Schematic of the acyl-transferase ribozyme where X= OH, NH_2. The substrate is a hexanucleotide (5′-CAACCA) bioconjugated to a biotinylated methionine at it's 3′-end. The ribozyme catalyses the transfer of the biotinyl-methionyl group to the 5′–hydroxyl or amino group, forming an ester or amide bond respectively[45].

However, the level of Amino–G incorporation only reached an observed maximum of 20 %, far lower than that of other novel nucleotides[45]. The cause of these disappointing results was thought to be either the low solubility of Amino–G in the aqueous transcription reaction mixture, or the presence of a cationic ammonium ion at its 5′–position. In an attempt to lower the proportion of Amino–G in the ammonium form, which was thought to take part in unfavourable interactions with the enzyme active site, the pH of the reaction was raised from the optimum conditions for T7 RNAP mediated transcription (from 8 → 9). The increase in pH level did not increase the level of incorporation above 20 %. This result in combination with the literature precedent for the use of novel nucleic acid initiators with either anionic or large neutral 5′–groups, demonstrating the flexibility of the T7 RNAP active site, indicated the solubility of Amino–G is likely to be a major factor in its ability to initiate transcription.

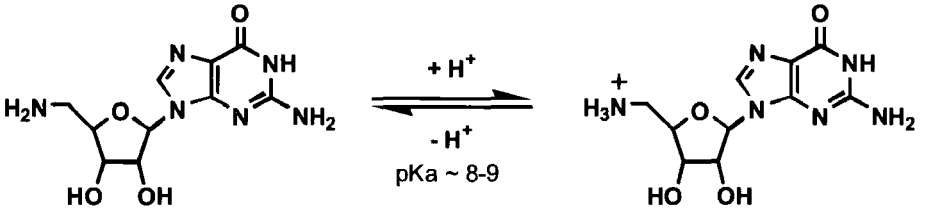


Fig:1.37 Base and conjugate acid forms of Amino–G.

1.34 Summary of novel nucleotide incorporation levels

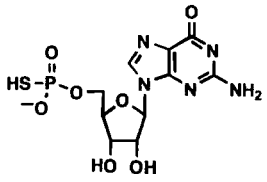
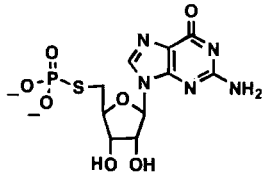
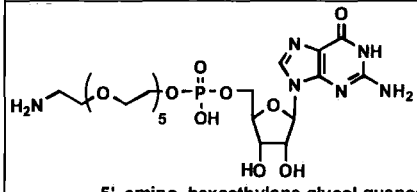
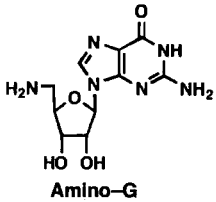
Novel Nucleotide	Level of incorporation (%)
 GMPS	80*[27]
 GSMP	80+[26]
 5'-amino-hexaethylene glycol guanosine (Amino-PEG)	65*[31]
 Amino-G	20*[45]

Table:1.3 Showing levels of incorporation of various novel nucleotides. * denotes the use of a reduced GTP concentration and a resulting lowered total RNA yield relative to a standard (0 mM novel nucleotide, 1:1:1:1 ratio of GTP:ATP:CTP:UTP).

1.35 Efficient incorporation of a terminal amine

Upon reviewing existing methodologies for the placing of an amine bioconjugation 'handle' at the 5'-terminus of RNA a number of observations can be made:

- Chemical modification (Section 1.19) is only applicable to small oligonucleotides and commonly used methods rely on the formation of an unstable phosphoramidate linkage.

- The Amino-PEG nucleotide can be incorporated into RNA chains at moderate levels, however the intricate synthetic route to the novel nucleotide could dissuade researchers with little experience of chemical synthesis[31].
- Synthesis of Amino-G is considerably easier, using few steps and no chromatographic techniques, however the low incorporation levels observed makes it a poor option.
- There is demand for an adaptable, quick and efficient method for incorporating amine groups at the 5′–terminus of RNA[31,45].
- If such a methodology could be developed it could provide an alternative to the unpleasant thiol techniques used as standard in many laboratories.

The following chapters detail the research we carried out in an attempt to develop such a new methodology for amine incorporation. As chemical synthesis and modification of RNA sequences has its limitations in both applicability and effectiveness, we aimed to develop an enzymatic route to amine incorporation. We wanted the technique to be as attractive and efficient as possible, therefore we put emphasis on a simple and direct synthetic route to a functional novel nucleotide. As a starting point we looked at Amino–G and hypothesised reasons for its low observed activity. From this standpoint we then defined a target molecule with a specific mode of action we theorised would provide increased levels of amine–incorporation.

2.0 Synthesis of GANP

2.1 Research Targets

Novel functional groups at the 5′-terminus of RNA provide valuable handles for bioconjugation to RNA molecules (see Section 1.18). Our aim is to follow a similar strategy to that used for the enzymatic incorporation of terminal sulfhydryl groups into RNA in order to devise a simple yet reliable method for the synthesis of 5′–AminoRNA. The modification of RNA in such a way will make available the option of utilizing amine chemistry in RNA studies where other techniques, such as sulfur chemistry, are unsuitable. For example, in a ribozyme selection experiment where the production of a free thiol could result in unwanted side reactions with the terminal sulfhydryl of GSMP primed RNA. The use of amine-primed RNA could circumvent the occurrence of these unwanted reactions and could be applied in roles similar to those currently occupied by thiol primed nucleic acids, such as investigating RNA-protein interactions, ribozyme selection/investigation, and linkage to other biomolecules. The use of 5′–AminoRNA would also be unaffected by the presence of oxygen, unlike the comparable thiol–RNA which is susceptible to oxidation.

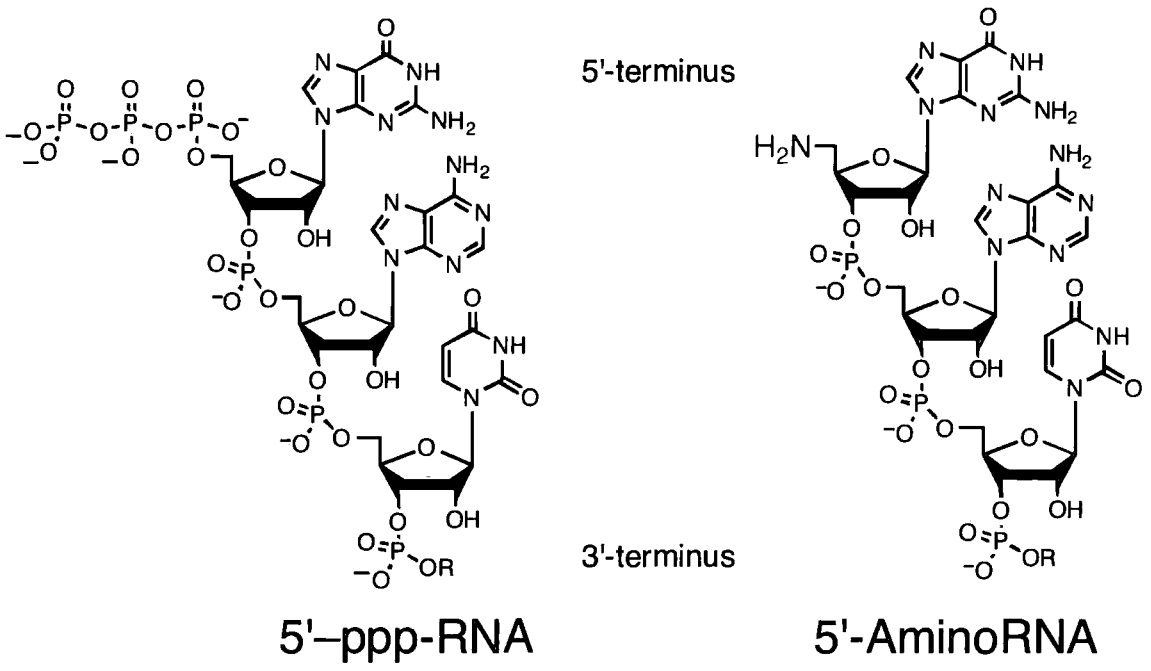


Fig: 2.1 Diagram of two RNA species; RNA featuring a relatively unreactive triphosphate 5′–group, and 5′–AminoRNA featuring a significantly more nucleophilic amine group at the 5′–terminus.

We chose to follow an enzymatic synthetic route over a chemical method owing to the success observed when using similar methodologies *i.e.* GMPS, GSMP[26-28,31]. Our intention is to produce an analogue of these novel nucleotides possessing a 5′-phosphoramidate rather than a phosphorothioate and assess its ability to initiate T7 RNA polymerase mediated RNA transcription. The subsequent hydrolysis of the phosphoramidate would reveal a 5′–AminoRNA species with the desired amine 'handle' at the 5'-terminus. RNA primed with a novel amino-nucleoside can be expected to undergo analogous conjugation reactions to those carried out by nucleic acids modified to incorporate terminal amines by other chemical or enzymatic methods *i.e.* conjugation to reporter/substrate molecules *via* amide bond formation.

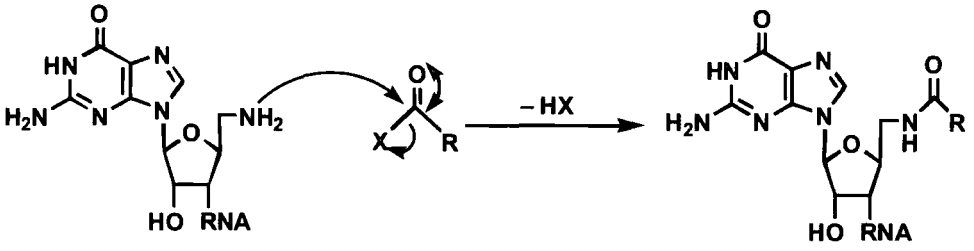


Fig:2.2 Bioconjugation to 5′–AminoRNA *via* a reaction with the incorporated nucleophilic amine, where **X** is a suitable leaving group, and **R** is a substrate/ reporter molecule.

The advantage of using an enzymatic method over chemical synthesis is primarily in the scale and applicability of the enzyme reaction. Unlike enzyme-mediated RNA synthesis chemical methods of oligonucleotide modification can only be applied to short RNA sequences. We aimed to improve on current technology for the formation of 5′-AminoRNA through the synthesis and use of 5'-amino-5'-deoxyguanosine 5'-N-phosphate (Fig:2.3), or as I shall refer to it from here on GANP, a novel guanosine nucleotide possessing a phosphoramidate group at the 5'–position.

5′–amino–5′–deoxyguanosine 5′–N–phosphate (GANP)

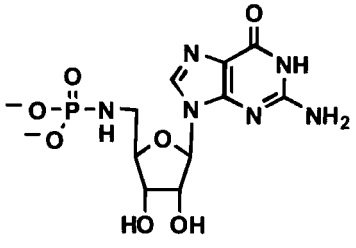


Fig:2.3 Dianionic form of GANP.

2.2 Theory of GANP action

Previously the use of Amino–G to initiate transcription of 5′-AminoRNA has yielded poor results, (see Section 1.33) the source of which was believed to be the low solubility of the amine in the aqueous transcription reaction[45]. We believed the introduction of an ionisable phosphoramidate group to the 5′–position would aid the solubility of the nucleotide and thereby improve both total RNA yield and the level of incorporation,

yielding 5′–phosphoramideRNA. The phosphoramidate group could then be hydrolysed to give a terminal amine group, thus producing 5′–AminoRNA.

As GANP is a mono– rather than a tri–phosphate nucleotide it does not possess the pyrophosphate leaving–group necessary to be incorporated into the body of the RNA transcript. This ensures phosphoramidate groups will only be present at the 5′–terminus of any resulting 5′–AminoRNA. The hydrolysis of the phosphoramidate groups incorporated into RNA transcripts can be carried out efficiently, owing to the pH dependent instability of such groups, to reveal a reactive amine group (Section 2.4). The amine will then be available for use as a bioconjugation 'handle'. As a comparison we have also investigated the ability of Amino–G to initiate T7 RNAP mediated transcription reactions.

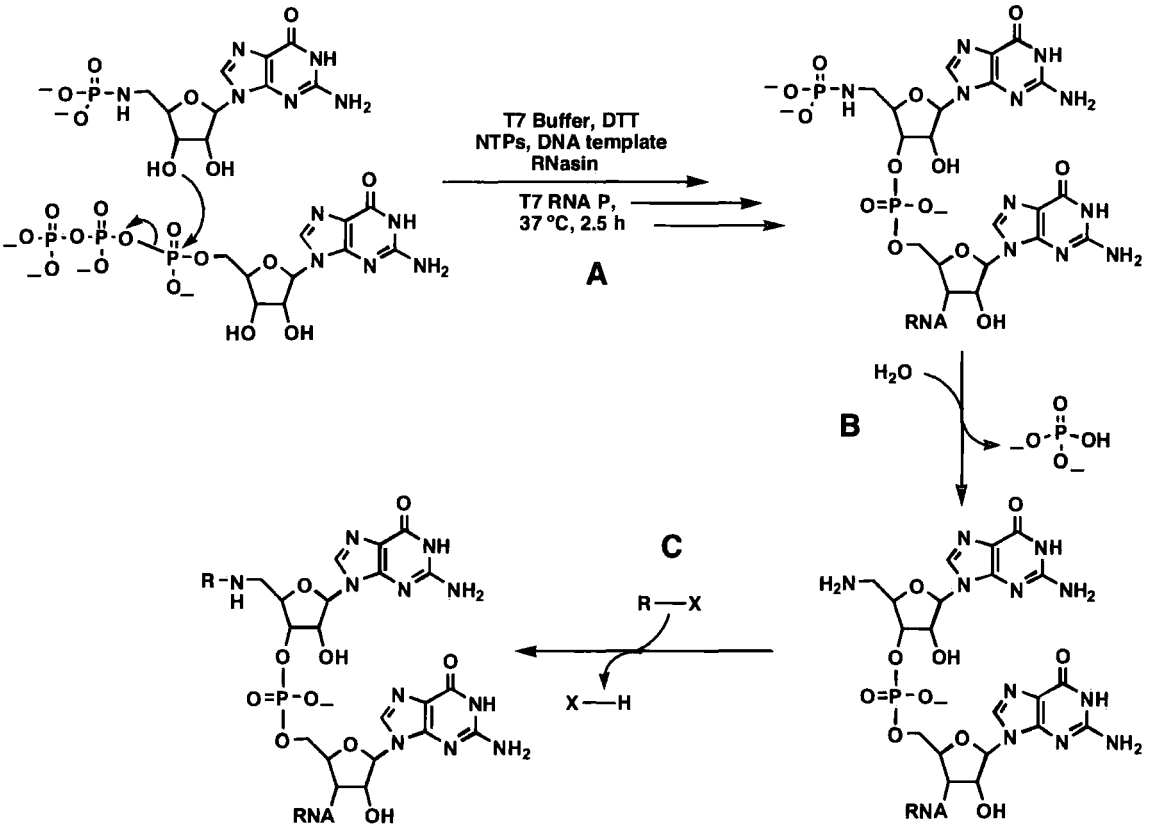


Fig:2.4 Planned role of GANP in: **A** acting as an initiator in transcription reactions, **B** hydrolysis of 5′–PhosphoramideRNA to reveal a reactive terminal amine group, and **C** specific bioconjugation of 5′–AminoG to desired functional groups.

We intended to devise a synthetic method using as few steps as possible, not only to maximise product yield but also to make the technology attractive to a broader range of researchers. By minimising the complexity of the route and avoiding specialised synthetic procedures we hope any methodology we develop would be useable not only by experienced chemists but other researchers with less synthetic experience.

2.3 Phosphoramidate chemistry

Our strategy for the synthesis and use of GANP is based on the properties of the phosphoramidate group, specifically that of their acid catalysed hydrolysis. The rate at which phosphoramidates hydrolyse to form the respective amine and phosphate will have a major impact on both the synthetic route we choose and how we handle our target novel nucleotide. The ability to alter the stability of phosphoramidates through pH control is the foundation to our strategy for successfully using GANP to initiate transcription reactions and the subsequent formation of 5′–AminoRNA.

The following sections will detail the results of a number of group's research into the hydrolysis of phosphoramidates and the development of theories used to explain their hydrolytic behaviour, beginning with early research and the development of a rate equation describing the contribution the various ionised forms of phosphoramidates make to the overall rate of hydrolysis. The early theory connected the pH dependent rate of hydrolysis to the pK_a of the ionisable groups on phosphoramidates, and in turn to the proportions of the ionic forms present at various pH levels[48]. As more research was undertaken an attempt was made to identify the nature of the monoanionic species' hydrolytic transition state[49]. Although the mechanism of hydrolysis proposed by the initial studies adequately explained the data collected, later research provided data in support of an alternative transition state. In the revised model correlations were made between the pK_a of the parent amine and the rate of phosphoramidate hydrolysis[50]. This in turn allowed approximations to the degree of bond breaking/ bond formation (β_{lg}, β_{nuc}) in the monoanionic hydrolysis transition state.

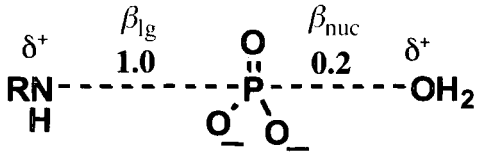


Fig:2.5 The proposed 'exploded' transition state of phosphoramidate monoanion hydrolysis[50].

2.4 Phosphoramidate pH rate profiles

In their 1953 paper J.D. Chanley and E. Feageson reported the results of their research into the hydrolysis of a number of aryl phosphoramidates in both water and dioxane:water solvent systems. Using the observed data they were able to both construct pH rate profiles for the hydrolysis of each of these compounds and form a general rate equation.

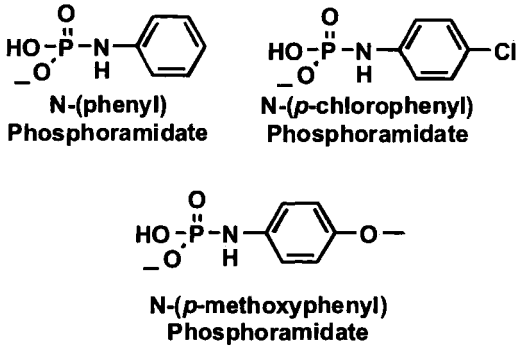


Fig: 2.6 The three phosphoramidate molecules used in Chanley and Feageson's research.

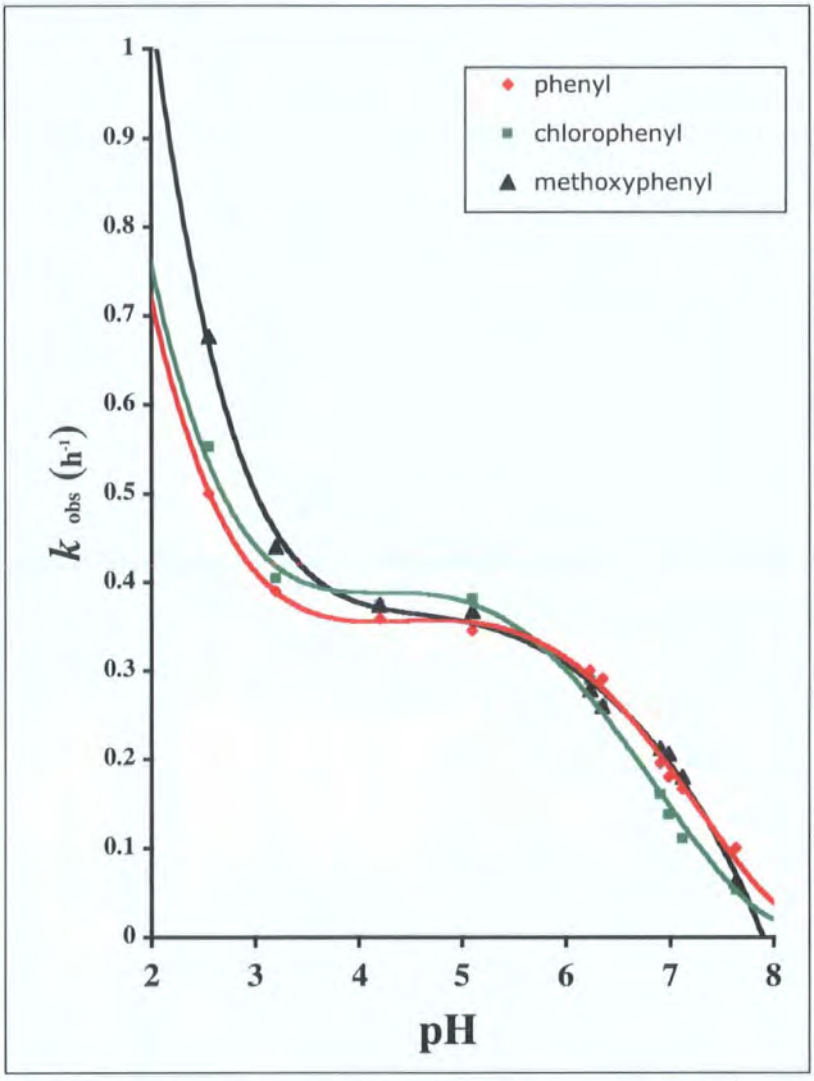


Fig:2.7 pH vs k_{obs} for a series of phosphoramidate species. The S-shaped curve followed by all examples is typical of phosphoramidate hydrolysis pH rate profiles.

The pH rate profiles for phosphoramidate hydrolyses can be broken down into three sections:

- Low pH (1 – 3) in which the fastest rates of hydrolysis are observed
- Mid–level pH (3 – 5) a pH independent plateau region with a moderate rate of hydrolysis ($t_{1/2} \sim 2$ h)
- High pH (5+) the region in which rate decreases as pH increases.

The pH levels at which transitions between these three regions occur were correlated to the pK_a values of the oxygen atoms in the phosphoramidate group by attributing the

different rates of hydrolysis in each section of the profile to the proportion of each ionic species present.

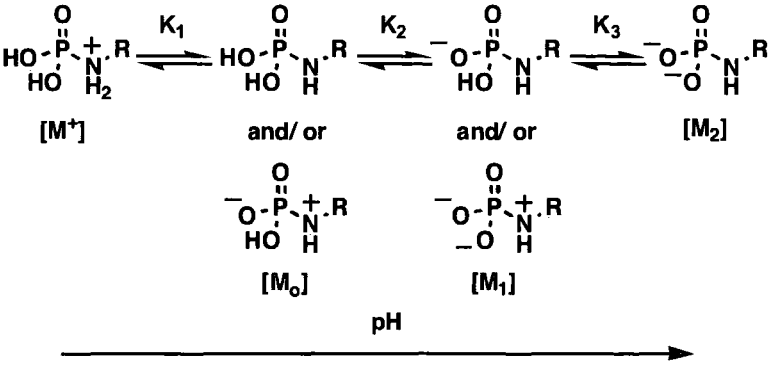


Fig:2.8 Ionic species of generic phosphoramidates.

2.5 Rate of hydrolysis at high pH

The pK_a of the second oxygen atom (pK_3) of phosphoramidates generally lies in the pH range 6.5-7.5[48]. As the pK_3 value is approached from low pH the proportion of fully ionised phosphoramidate species ($[M_2]$) in solution increases. Chanley and Feageson asserted that the dianion is completely inert to hydrolysis, possibly owing to the electron density on the oxyanions repelling nucleophilic species[48,49]. Hence, as the mole fraction of $[M_2]$ in solution increases, the rate of hydrolysis decreases. As the dianion $[M_2]$ does not undergo hydrolysis at an appreciable rate it does not have a term in the rate equation[48,49].

2.6 Rate of hydrolysis in plateau region

Chanley and Feageson defined the plateau region as lying between the pK_a values of the two oxygen atoms of the phosphoramidate (pK_2 and pK_3). Therefore the $[M_1]$ species is the most prevalent in the plateau region, with almost 100 % of the phosphoramidate present in either the zwitterionic or mono-anionic form. Over a 100-fold change in hydrogen ion concentration the rate of hydrolysis remains constant, the invariance in rate over this range supporting this claim. As only the monoion species ($[M_1]$) is present in the majority of this pH range, rate measurements can be used to determine the associated

specific rate constant, k_1. The pK_3 value of the phosphoramidate can also be estimated from the rate vs pH curve by noting the pH at which rate is half that of the flat region *i.e.* where 50 % of $[M_1]$ is deprotonated to form $[M_0]$.

2.7 Rate of hydrolysis at low pH

At pH levels below pK_2 of the phosphoramidate, *i.e.* at pH below the plateau region, the rate of phosphoramidate hydrolysis in water increases sharply. This observation has been attributed to an increase in the concentration of $[M_0]$ and $[M_1]$ species, which in turn leads to an increase in the pseudo first order observed rate constant. However Chanley and Feageson believed that the increase in the value of rate constant as the pH lowers is slower than that expected were acid catalysis the only hydrolysis mechanism occurring, therefore they postulated an additional decomposition pathway involving the build up and hydrolysis of an uncharged species $[M_o]$.

2.8 Hydrolysis rate equation

Through correlation of each section of the pH rate profile to distinct ionic species Chanley and Feageson found it possible to define a rate equation for the hydrolysis, with a rate constant attributed to each ionic form.

$$\text{Rate} = k_{obs}\,[S]$$

Where under pseudo 1st order conditions:-

$$k_{obs} = k_H[H^+][M_o] + k_o[M_o] + k_1[M_1] \quad [48]$$

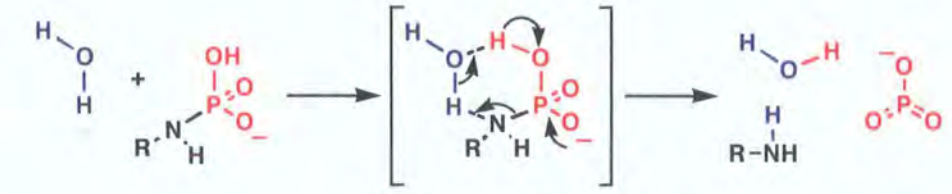
Where M_o and M_1 are the concentrations of the neutral and singly negatively charged species respectively, and k_H, k_o and k_1 are the associated specific rate constants. The dianionic species $[M_2]$ has no term in the rate equation, as the rate of its hydrolysis is so low as to be negligible to the overall reaction rate. There is no term for the charged ion $[M^+]$ as its hydrolysis and that of $[M_o]$ in the presence of acid are indistinguishable. As the rate constants are each related to a particular phosphoramidate species it is possible to determine their values by observing hydrolysis reactions under conditions where only one

of those species is present *e.g.* the value of k_1 can be evaluated by observing the rate of phosphoramidate hydrolysis within the plateau region.

2.9 The mechanism of hydrolysis

The mechanism of hydrolysis of the uncharged phosphoramidate species (corresponding to the term $k_0[M_0]$) was suggested by Chanley and Feageson to function *via* the unimolecular degradation of a zwitterionic transition state. The hydrolysis of the monoionic phosphoramidate species was known to be acid catalysed though the mechanism of reaction was unknown. Hydrolysis of the phosphoramidate is known to result in the production of an amine and a phosphate species, the nature of which is dependent on the reacting nucleophile, when water acts as the nucleophile, phosphoric acid is formed, when an alcohol reacts a phosphate monoester is produced. What is unknown is whether or not the mechanism of phosphate production operates *via* an extremely reactive metaphosphate species, as is thought to occur in the hydrolysis of phosphate esters or *via* a biomolecular mechanism in which no metaphosphate is formed. Metaphosphate reacts at such high rates with nucleophilies that its identity cannot be confirmed by spectroscopic methods.

Two possible mechanisms for the hydrolysis of the monoanion species ($[M_1]$) were proposed by Chanley and Feageson, both of which feature a proton transfer to the amine group, and result in the formation of a metaphosphate species.



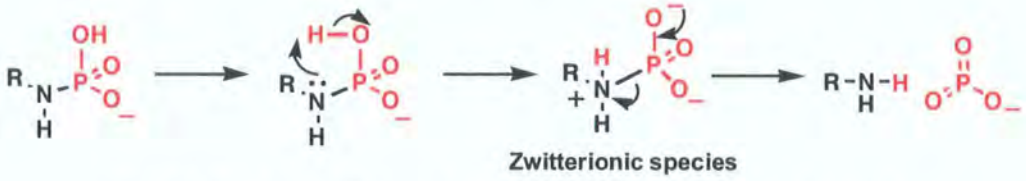


Fig: 2.9 Chanley and Feageson's proposed mechanisms of phosphoramidate monoanion hydrolysis. The hypothesised 'Shift' mechanism features a bimolecular cyclic transition state that incorporates a mono-anionic phosphoramidate species and a water molecule. The zwitterion mechanism was proposed to operate *via* an internal proton transfer producing the zwitterionic form of the phosphoramidate, which then undergoes unimolecular degradation in the transition state to form a terminal amine and a metaphosphate species[48].

Collapse of the cyclic transition state achieves cleavage of the P-N bond and simultaneous proton transfer. No zwitterion is formed and metaphosphate is expelled[49]. This mechanism is more favourable in phosphoramidates than it would be in phosphate esters, which are hypothesised to undergo an intramolecular proton transfer owing to the basic nature of the nitrogen atom compared the oxygen of the ester, making possible the abstraction of a proton from a solvent molecule.

A hydrolysis mechanism that proceeds by an internal proton transfer and zwitterion formation does not wholly agree with experimental results because hydrolysis reactions proceeding *via* such a mechanism would be expected to display a decrease in k_{obs} owing to the unfavourable formation of a zwitterionic species in a less polar solvent system. The decrease in rate observed for the hydrolysis of N-(*p*-chlorophenyl) phosphoramidate 1:1 dioxane/ H_2O conditions (see Fig:2.10) was determined by Chanley and Feageson to be too low to corroborate a zwitterionic transition state pathway. Although a reduction in rate is evident in the mid pH range the decrease was judged too small to support the

occurrence of an internal proton transfer. Therefore the 'Shift' mechanism with its cyclic transition state was put forward as the most likely mechanism.

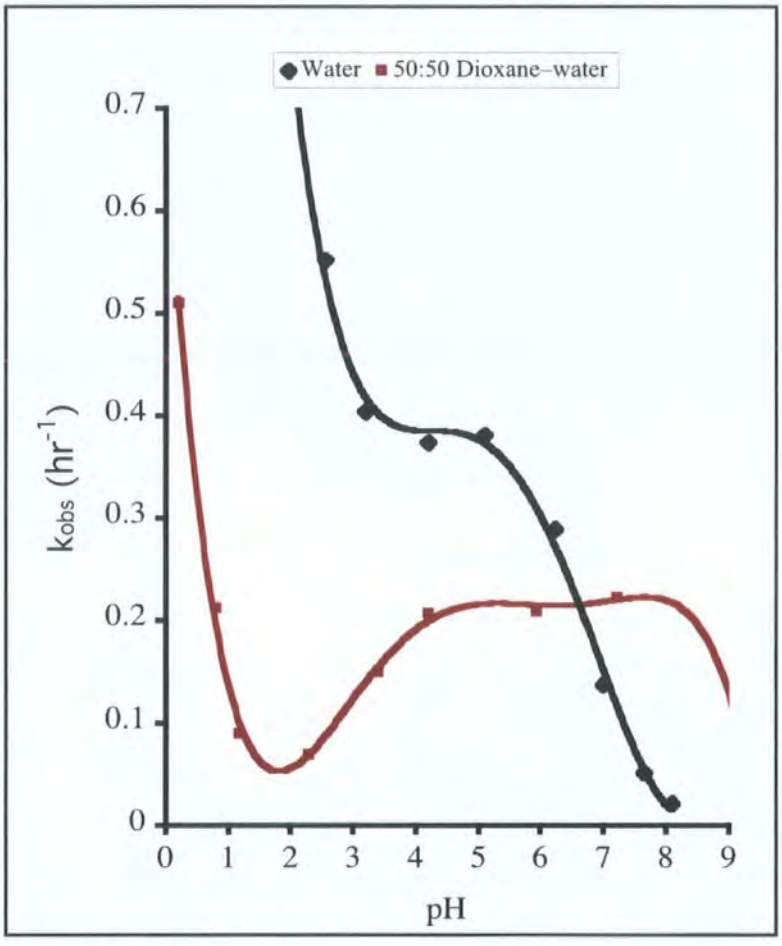


Fig: 2.10 Plot of the experimental rate constant vs pH for the hydrolysis of N-(*p*-chlorophenyl) phosphoramidate in both aqueous and aqueous:dioxane solutions. The decrease in rate evident in the mid pH region was judged to be insufficient to support the presence of a zwitterionic transition state[48].

2.10 Evidence for a metaphosphate intermediate

Both the zwitterionic and 'Shift' mechanisms hypothesised by Chanley and Feageson produce the short–lived metaphosphate species, however no evidence confirming the presence of this elusive group had been presented. To provide proof for the production of metaphosphate during phosphoramidate hydrolysis and thereby support their hypothesised reaction mechanism, Chanely and Feageson undertook further research to

collect data of the product distribution resulting from hydrolysis of phosphoramidates in solvents of various ratios of water: alcohol[49]. The introduction of an alcohol into the reaction mixture makes possible a solvolysis mechanism whereby the alcohol reacts with the phosphorus group to form the respective alkyl-phosphate, rather than the inorganic phosphate formed through reaction with water (see Fig:2.11). If the hydrolysis mechanism proceeds *via* the formation of a metaphosphate species it would be expected that the proportion of phosphorus of containing products (phosphate and alkyl-phosphate) would the same as the mole ratio of water: alcohol present in the reaction mixture. This product distribution would occur, as the reactivity of the metaphosphate species is so high that the selectivity of phosphates produced would be determined not by the nucleophilicity of the alcohol and water but by which species the metaphosphate collided with first.

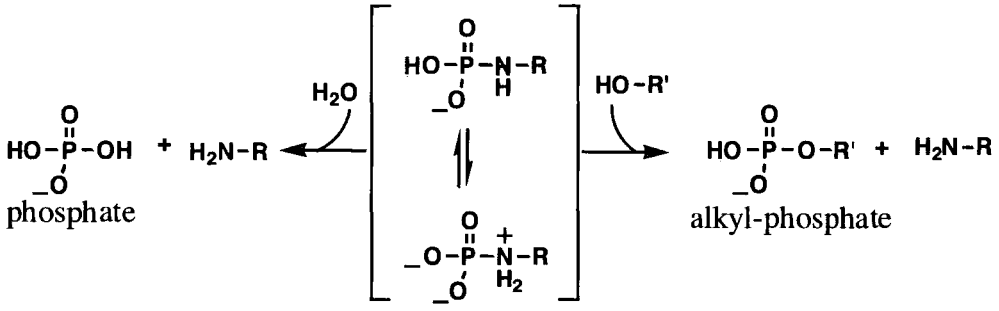


Fig:2.11 Reaction scheme showing solvolysis products of a phosphoramidate monoanion when reacting with water and an alcohol respectively.

If the hydrolysis operates *via* nucleophilic attack at the phosphorus atom, with no metaphosphate intermediate being formed, the ratio of products would differ from the mole ratios of solvents, owing to the higher nucleophilicity of methanol.

2.11 Results of solvolysis experiments

Unfortunately the results of this study were not as clear as the researchers may have hoped, though some conclusions were drawn as to the nature of the reaction mechanisms taking place.

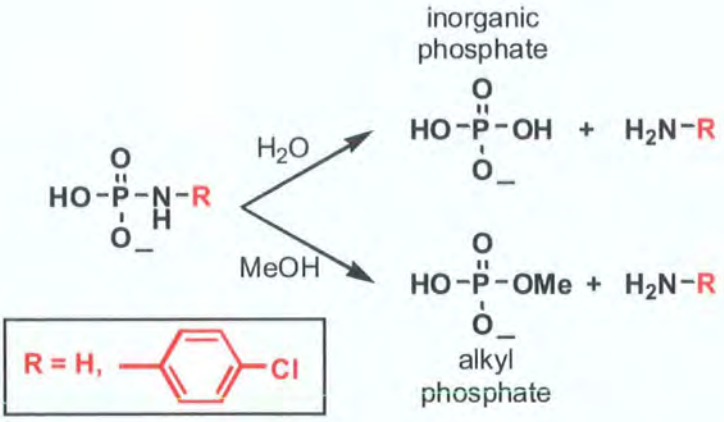


Fig:2.12 Products of phosphoramidate solvolysis with water and methanol.

	% alkyl-phosphate produced	
mol % Methanol	**R = H**	**R= *p*-chlorophenyl**
9.8	38.8	–
15.5	55	22
29.2	61.2	38
48	–	55

Table: 2.1 Data above gives % of alkyl phosphate produced in experiments following the hydrolysis/solvolysis of Phosphoramidic acid (R=H) and (*p*-chlorophenyl) phosphoramidate (R= *p*-chlorophenyl as in the reaction scheme shown in Fig:2.10. The pH of each reaction was adjusted to 5, thus ensuring the monoanion was the most prevalent ionic species[49].

The degradation of both phosphoamidic acid, and (*p*-chlorophenyl) phosphoramidate was observed. The percentage of alkyl-phosphate the phosphoramidic acid produced during solvolysis was far above the mole percent of methanol used in each experiment (see Table:2.1), indicating a reaction mechanism with a bimolecular rate–determining step where no metaphosphate is produced and nucleophilicity of the alcohol is a key factor. The product distribution observed for *p*–chlorophenyl phosphoramidate were closer to those expected if a unimolecular rate-determining step with metaphosphate production took place, *i.e.* a proportion of alkyl-phosphate close to the mole percent of alcohol used was observed. As the products of reaction differ between the two substrates, it is clear the reaction mechanisms taking place are not equivalent and that more complex models than the 'Shift' mechanism would be required to describe the observed data.

2.12 Conclusions from mixed solvent experiments

As a model system Chanely and Feageson looked at the degradation of O–phosphate esters in mixed methanol:water solutions, which produce monomethyl phosphate and phosphoric acid. In this case the product distribution observed is approximately equal to the mole ratio of the solvent mixture, as was the case with (*p*-chlorophenyl) phosphoramidate. These observations were used to support the formation of a metaphosphate species upon O–phosphoester hydrolysis.

Owing to the structural similarity between O–phosphoesters and phosphoramidates Chanley and Feageson proposed a similar mechanism for the solvolysis of phosphoramidic acid. However, the results shown in Table:2.1 infer that this assumption was incorrect, and that a bimolecular process involving direct attack of a solvent on the phosphorus provides a better fit with the experimental data *i.e.* a product distribution of inorganic phosphate and monophosphate ester at odds with the mole proportions of water and methanol in the reaction mixture. For example in a solution of 15 mol % methanol the solvolysis of phosphoramidic acid resulted in a product distribution of 55 % monophosphate ester, indicating that the nucleophilicity of the alcohol had a significant effect on the selectivity of the reaction.

Based on the observation that the O–phosphoester model was a poor fit for the hydrolysis of phosphoramidates, Chanely and Feagson formulated a rate equation for the degradation of the monoanionic species of the phosphoramidic acid [M_1] in mixed alcohol:water solvents:

$$k_{obs} = k_{water}(H_2O) + k_{methanol}(CH_3OH) \quad [49]$$

The derivation of k_{water} and $k_{methanol}$ from experimental product distribution data allowed Chanley and Feageson to predict what proportion of phosphoramidic acid would react with methanol to form alkyl phosphate in various solvent mixes (see Table:2.2).

	% alkyl phosphate	
Mole % MeOH	**Observed**	**Calculated**
4.7	23.1	25.6
9.8	38.8	43.4
15.5	55	54.9
22	61.2	66.4
29.3	72.5	74.4

Table:2.2 Comparison of calculated and experimental data for the solvolysis of Phosphoramidic acid monoanion in a range of solvent mixes. The pH used in each experiment ensured ~99% of the acid was present as the monoanion[49].

A close fit between the calculated and experimental data strongly indicated the bimolecular reaction scheme on which the rate equation was based adequately described the solvolysis mechanism for phosphoramidic acid. The solvolysis of (*p*-chlorophenyl) phosphoramidate was also presumed to function *via* a bimolecular pathway, although the product distribution of solvolysis reactions (Table:2.1) indicated this route was not equivalent to that of phosphoramidic acid.

2.13 The two mechanisms of solvolysis

The disparity observed between the results for phosphoramidic acid and *p*–chlorophenyl phosphoramidate was qualitatively explained by considering the equilibrium between the two forms of the monoanion [M_1], the non-zwitterion and zwitterion.

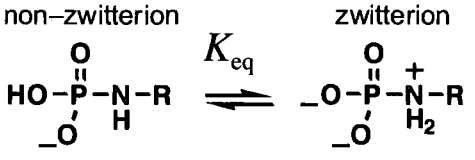


Fig:2.13 Equilibrium between the two forms of the phosphoramidate monanion, the non-zwitterion and zwitterion.

The value of K_{eq} in water for the phosphoramidic acid was presumed by Chanley and Feageson to be large, with the majority of monoanion in the zwitterionic form, an assumption corroborated by crystallographic data that revealed that the acid is a zwitterion in the solid state. Therefore as the addition of methanol reduced the polarity of the solvent the position of the equilibrium would not be substantially altered towards the

left hand side. Conversely Chanley and Feageson presumed the K_{eq} value for (*p*-chlorophenyl) phosphoramidate in water was small, and hence the addition of methanol would have a more noticeable effect and shift the equilibrium to the left, increasing the proportion of the non–reactive, non–zwitterionic, form present in solution.

To explain the difference in product distribution observed experimentally, separate mechanisms were assigned to the solvolysis of the non–zwitterion and zwitterionic forms (see Fig:2.14). The relative stability of the neutral species (non-zwitterion) and diionic species of phosphoramides, as well as that of the anion of diamidophosphoramides $(RNH)_2PO_2^-$ suggests that not only a protonation of the nitrogen is involved in the solvolysis, but that this proton is derived from the P-OH moiety. Chanley and Feageson determined that the magnitude of the observed deuterium–isotope effect was insufficient to indicate that a proton transfer is involved in the rate-determining step.

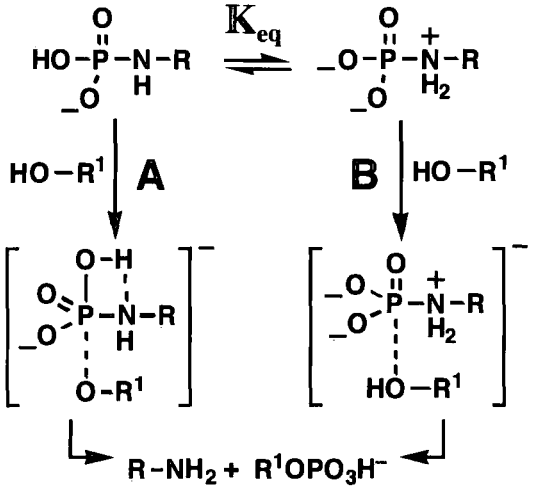


Fig:2.14 Mechanisms for the solvolysis of **A** the non-zwitterion and **B** the zwitterion. Both routes result in the same products, however Chanley and Feageson expected the rate of route **B** to show a dependence on the nucleophilicity of the incoming species[49].

Of the two mechanisms shown in Fig:2.14 Chanley and Feageson believed route **B** to be the most prevalent as it best explained the observed product distributions for phosphoamidic acid (Tables:2.1–2.2). However, the alternative route (**A**) was also deemed important under set conditions, *i.e.* in phosphoramidates where the equilibrium

constant K_{eq} (in water) is very small, the introduction of a less polar solvent can cause route **A** to be favoured owing to the increased proportion of non-zwitterion present. The presence of two separate mechanisms for the zwitterionic and non-zwitterionic forms of the monoanion was also applied to the two forms of the neutral species (see Fig:2.15), with the provision that the non-zwitterion was considered to be far less reactive than its zwitterion counterpart.

Fig:2.15 The two forms of neutral phosphoramidates, the non-zwitterion (left) and zwitterion (right) species. The high rate of phosphoramidate solvolysis (see Fig:2.7) at low pH was attributed to the instability of the zwitterionic species.

However, Chanley and Feageson provided no experimental evidence to conclusively support the hypothesis that separate mechanisms exist for the solvolysis of the zwitterionic and non–zwitterionic species, leaving room for further research into the nature of the hypothesised reaction mechanisms and transition states.

2.14 Revised mixed solvent data

In 1971 Benkovic and Sampson published a paper in the Journal of the American Chemical Society entitled 'Structure-Reactivity Correlation for the Hydrolysis of Phosphoramidate Monoanions' in which they provided experimental evidence and a plausible theory for linking the rate of phosphoramidate solvolysis to the pK_a of the amine leaving-group[50]. To give a full account of how they came to this conclusion it is pertinent to first cover their treatment of product distributions in the solvolysis of a range of phosphoramidate monoanions in mixed solvent solutions such as those reported in Chanley and Feagesons research covered in Sections 2.10–2.13. I will then go on to discuss how their research developed, resulting in their proposed structure for the transition state of phosphoramidate monoanion solvolysis.

Phosphoramidate	Solvent	% Methyl Phosphate	$k_{obs} \times 10^3$ (min^{-1})
N-(*n*-Butyl)	H_2O		1.64
	D_2O		1.51
	CH_3OH–H_2O (50 v/v)	74	
N-Methylimidazole	CH_3OH–H_2O (50 v/v)	78	
N-(*p*-Carboxyphenyl)	CH_3OH–H_2O (50 v/v)	71	
Phosphoamidic acid	H_2O		4.02
	D_2O		3.48
	CH_3OH–H_2O (50 v/v)	73	4.71
	Dioxane–H_2O (50 v/v)		5.25
N-(Chlorophenyl)	H_2O		11.5
	D_2O		15.2
	CH_3OH–H_2O (50 v/v)	68	8.33
	Dioxane–H_2O (50 v/v)		6.73

Table:2.3 Solvolysis of phosphoramidate monoanions[50].

The researchers four main observations were:

1. The product distribution in mixed methanol-water solvent systems leads to a preponderance of methyl phosphate over inorganic phosphate.
2. The selectivity for methanol to water is essentially independent of the phosphoramidate used (~70 % at CH_3OH–H_2O)[49].
3. The observed rate constant of solvolysis of phosphoramidic acid is increased in less polar solvents relative to purely aqueous media while the rates of solvolysis of *N*-(*p*-carboxyphenyl and *N*-(*p*-chlorophenyl)phosphoramidates are decreased.
4. There was no significant kinetic isotope effect observed when using deuterium oxide in place of water ($k_D/k_H \sim 0.9$).

The results and subsequent observations fit in part with the hypothesis detailed in Section 2.13, *i.e.* no deuterium induced kinetic isotope effect, therefore no proton transfer in the rate- determining step. However, the product distribution of solvolysis in mixed solvent systems was found to be independent of the nature of the phosphoramidate, which was at odds with the results of Chanely and Feagson. The significance of this conclusion and its effect in determining a mechanism and transition state for solvolysis will be discussed below, after first introducing another set of significant experimental results from Benkovic and Sampson's work.

2.15 Correlation of k_{obs} to pK_a

The authors brought together kinetic data for the hydrolysis of a wide variety of phosphoramidates (Fig:2.16, Table:2.4). The phosphoramidates were organised according to the nature of their parent amines into classes I (alkyl amines), II (pyridinyl derivatives), III (mostly aniline derivatives) and a correlation plot between the k_{obs} for the hydrolysis of each monoanion and the pK_a of the conjugate acid of the corresponding parent amine was constructed.

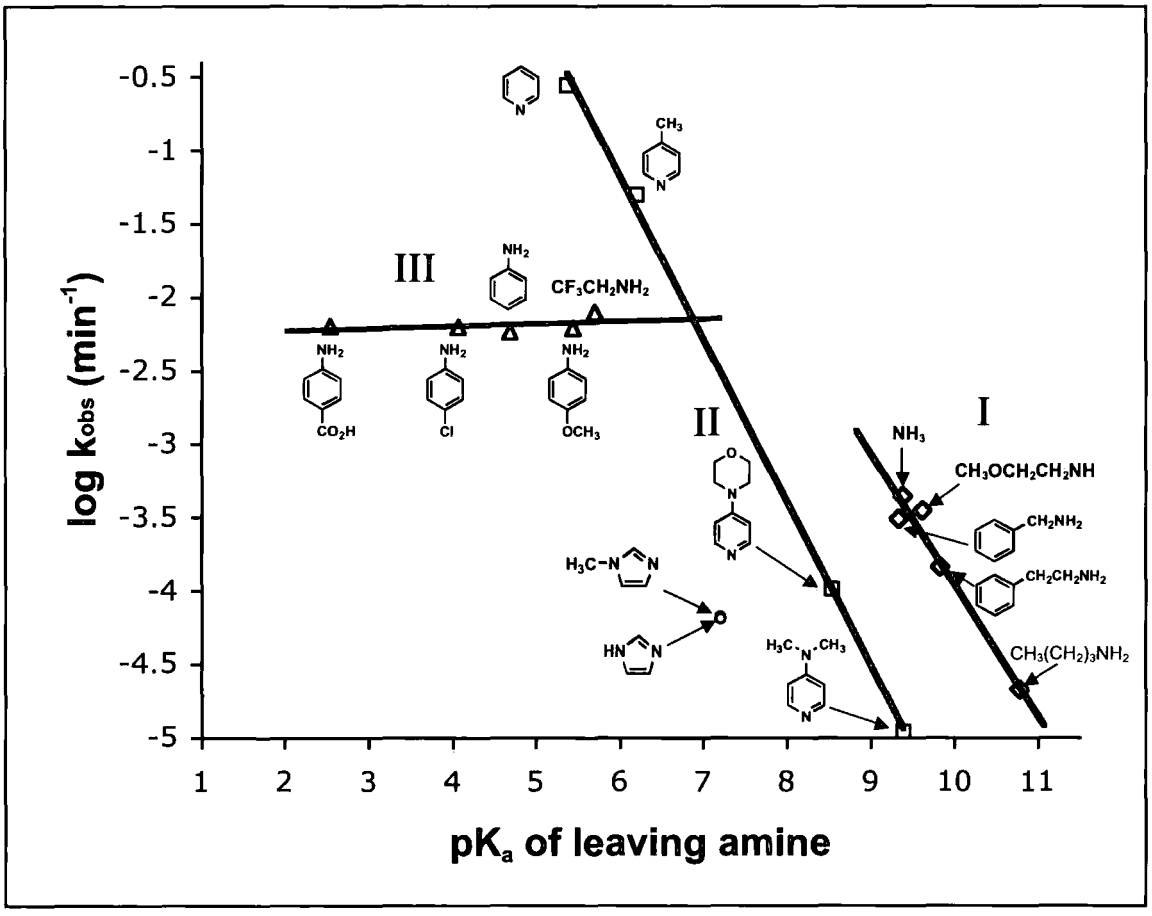


Fig:2.16 Plot of log k_{obs} (min^{-1}) for the hydrolysis of phosphoramidate monoanions at 20 °C *vs.* the pK_a of the conjugate acid of the leaving amine taken from Benkovic and Sampson[50].

Parent Amine	pK_a of parent amine	$k_{obs} \times 10^4$ (min^{-1})
p-Carboxyphenyl	2.54	63.1
p-Chlorophenyl	4.07	62.8
Phenyl	4.69	58.3
p-Methoxyphenyl	5.44	61.7
2,2,2-Trifluoroethyl	5.7	79.2
Pyridine	5.37	2780
4-Methylpyridine	6.21	497
4-Morpholinopyridine	8.53	1.02
4-Dimethylaminopyridine	9.39	0.11
N-Methylimidazole	7.2	0.67
Imidazole	7.2	0.65
Ammonia	9.38	4.42
Methoxyethyl	9.61	3.52
Benzyl	9.33	3.07
Phenethyl	9.83	1.46
n-Butyl	10.77	0.212

Table:2.4 Rate coefficients for the hydrolysis of Phosphoramidates at 20 °C[50].

Of the three classes, the k_{obs} values of classes I and II showed high sensitivities to pK_a (β = 1.0), increasing as pK_a decreases. Contrary to this the class III phosphoramidate k_{obs} values proved invariant to the pK_a of the leaving group. In an attempt to justify both these observations and those made from the mixed–solvent hydrolysis experiments (see Section 2.14) within a common reaction mechanism comparisons were first drawn to the well–defined hydrolysis of phosphate esters, which is the subject of the next section.

2.16 Mechanism of phosphate monoester hydroylsis

The k_{obs} for the hydrolysis of the monoanions of phosphate monoesters have been found to fit the equation:

$$\log k_{obs} = 0.91 - 0.27pK_a$$

Where the pK_a is that of the parent alcohol.

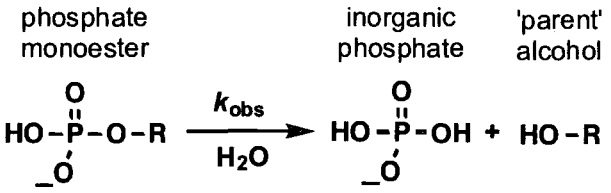


Fig:2.17 Hydrolysis of phosphate monoesters produces inorganic phosphate and an alcohol.

This relationship has been found to hold true over a pK_a range of approximately 12 units. Benkovic and Sampson hypothesised that the mechanism of hydrolysis featured a pre-reaction equilibrium between the zwitterion and non-zwitterionic forms of the monoanion. They proposed that only the zwitterion is able to undergo a rate determining unimolecular decomposition to form the alcohol and metaphosphate products. The metaphosphate undergoes an extremely fast reaction with any available nucleophile to complete the reaction scheme (Fig: 2.18).

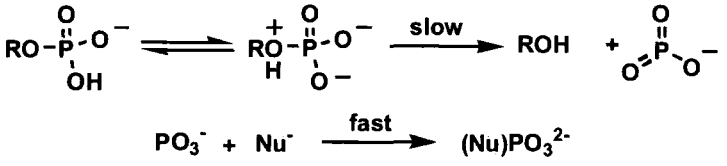
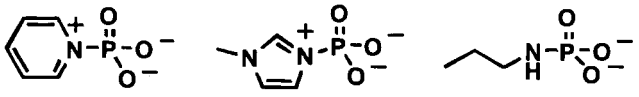


Fig:2.18 Hypothesised mechanism of monoester phosphate (monanion) hydrolysis.

Benkovic and Sampson proposed that a similar mechanism operated in the hydrolysis of phosphoramidate monoanions, a hypothesis supported by:

- experimental evidence:- phosphorylpyridinium and phosphoryl-N-methylimidazolium, which possess a cationic N independent of pH, retain reactivity under high pH conditions where phosphoramidaties derived from primary amines exist as unreactive dianions (see Fig:2.18).
- Theoretical evidence:- decomposition of the zwitterionic form produces a neutral amine leaving group rather than a highly unfavourable anionic amine.



(1) (2) (3)

Fig:2.19 Phosphorylpyridinium (**1**) and phosphoryl-N-methylimidazolium (**2**) molecules retain their zwitterionic character and hydrolytic behaviour at high pH levels. Primary amine derived phosphoramidates (**3**) however exist as dianions, and display very low levels of reactivity.

2.17 Zwitterion formation in Class I + III phosphoramidates

The formation of a zwitterion in class I and III phosphoramidates is inexorably linked to the pK_2 of the molecule, specifically whether this pK_a value relates to the protonation of the orthophosphate oxyanion or the nitrogen atom (Fig:2.20–21). If the pK_a of the conjugate acid of the nitrogen atom is higher than that of the orthophosphate ($pK_{orthophosphate}$ = 7.2) then as pH is reduced from >pK_2 protonation of nitrogen will be favoured over the oxygen. In such a situation the zwitterionic form will dominate over the non-zwitterionic form in aqueous solution. If the pK_a of the nitrogen is lower than that of the orthophosphate however, the opposite holds true, and the non-zwitterionic form dominates. It is this key factor that was hypothesised to result in the sensitivity of Class I phosphoramidates to the pK_a of the parent amine, while the Class III compounds displayed a marked insensitivity.

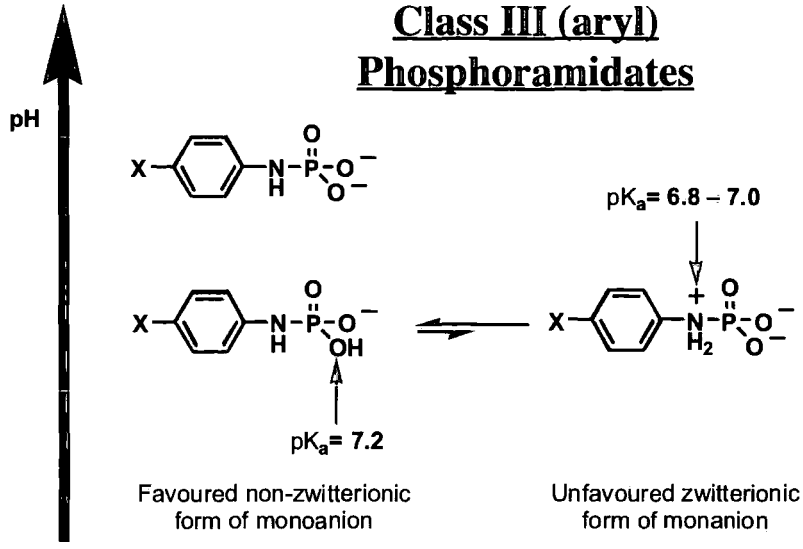


Fig:2.20 Schematic showing the protonation of Class III phosphoramidates as pH is reduced, as determined by the pK_a values of the nitrogen and orthophosphate groups. As shown this results in the non-zwitterionic form of the monoanion being favoured over the zwitterion form[50].

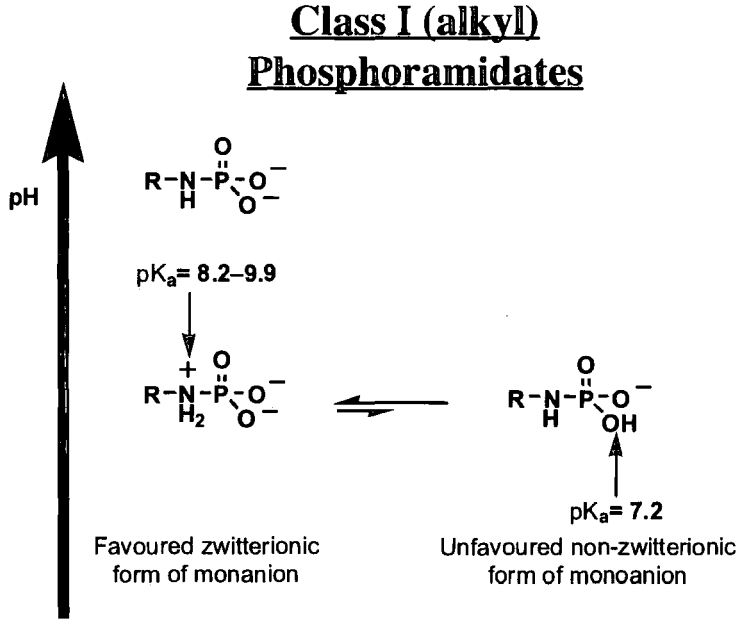


Fig:2.21 Schematic showing the protonation of Class I phosphoramidates as pH is reduced, as determined by the pK_a values of the nitrogen and orthophosphate groups. Contrary to Class III phosphoramidates the pK_a of the nitrogen is higher than that of the oxygen, the result of which is a favoured zwitterionic species[50].

Plots of pK_2 *v s* pK_{amine} of the conjugate acid of the parent amine for Class I phosphoramidates was found to have a gradient of –1, indicating the pK_2 of this class of phosphoramidate are directly proportional to the pK_{amine}. This relationship indicates that the attachment of a phosphoryl group to the amine group results in the pK_{amine} lowering by 1 unit to give the pK_2 of the phosphoramidate. From these observations Benkovic and Sampson were able to make a semi-quantitative estimate of the equilibrium–constant (K_{ZW}) for the distribution of the monoanionic species of Class I phosphoramidates:

$$K_{zw} = \frac{[\text{ zwitterion }]}{[\text{ non-zwitterion }]}$$

$$\log K_{zw} = pK_2 - 7.2 \quad [50]$$

By definition a pK_2 of 7.2 gives a K_{ZW} value of 1, i.e the orthophosphate and nitrogen atoms are protonated in equal abundance, leading to equal concentrations of both zwitterionic and non-zwitterionic species. This relationship cannot be applied to the Class III as the pK_2 of those phosphoramidates is essentially that of the orthophosphate, and owing to the minimal transmission of electronic effects through the phosphorus atom, almost completely invariant to change in the parent amine.

2.18 Rate equation for monoanion hydrolysis

Assuming hydrolysis of phosphoramidate monoanions follows a mechanism similar to that of phosphate monoesters (see Fig:2.18) a rapid pre–reaction equilibrium exists for the formation of the zwitterion. As the zwitterion is the reactive form of the monoanion this equilibrium must therefore feature in the rate equation for the hydrolysis:

$$k_{obs} = k_r \frac{K_{zw}}{(K_{zw} + 1)}$$

where:

k_r = rate of hydrolysis of the monoanionic zwitterion

$\frac{K_{zw}}{(K_{zw} + 1)}$ = mole fraction of zwitterion

In Class I where:

$$K_{zw} >> 1$$

$$k_{obs} = k_r$$

So that k_r depends on the pK_a of the leaving amine with a slope (β) of –1.00.

In Class III:

$$K_{zw} << 1$$

$$k_{obs} = K_{zw} k_r$$

The sum of the Brønsted slope for k_r (-1.00) and the dependency of K_{zw} on the pK_a of the amine (+1.00) equals zero. A consequence of this is that the observed rate remains relatively constant across the class as pK_a is varied. In agreement with this theory the rate of phosphoryl pyridinium hydrolysis is 40 times that of N-(*p*-methoxyphenyl) phosphoramidate, even though the parent amine of both species have a pK_a of 5.4.

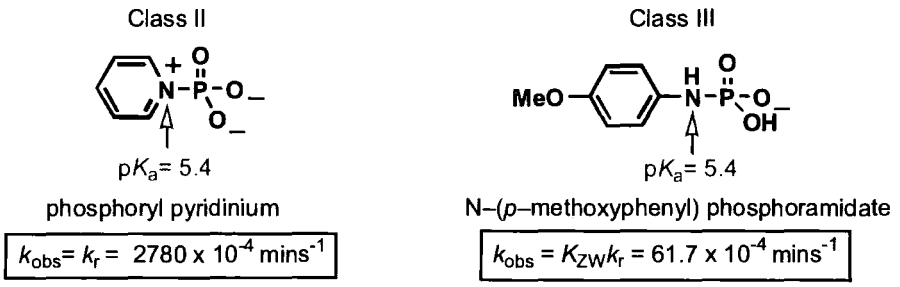


Fig:2.22 Illustration of the differing rates of reaction of zwitterionic and non–zwitterionic species. The pK_a values shown are those of the conjugate acids of the parent amines.

2.19 Solvent effects

As shown in Table:2.4 the media in which hydrolysis/solvolysis occurs can have a marked effect on the rate of phosphoramidate decomposition. What is also clear is that the effect is variable across the different Classes of phosphoramidate. By using the correlation between pK_2 and reaction rate set out above, Benkovic and Sampson formed a hypothesis to explain the variation in the observed rate constant through the different classes.

In Class III phosphoramidates the change in solvent from an aqueous media to an alcohol–water mix results in a decrease in the value of k_{obs} for the hydrolysis/solvolysis. This change occurs because of the effect the less polar media has on the formation of the already unfavourable zwitterionic species. A similar effect has been reported in aliphatic amino acids where:

$$K_{zw} = 10^5 \quad \text{in } H_2O$$

$$\downarrow$$

$$= 1 \quad \text{in EtOH}$$

In a less polar solvent system the formation of charged molecules is less favourable as the creation of a supporting electrostatic solvent sphere is less likely. Although the formation of an ordered solvation system has an entropic cost, this is more than balanced

by the enthalpic bonus of the electrostatic interactions between zwitterion and solvent molecules. Hence, a reduction in K_{ZW} is observed in a non-polar solvent. For Class III phosphoramidates this reduction in the proportion of monoanion in the zwitterionic form has greater implications than for other Classes, as the value of k_{obs} is directly proportional to K_{ZW}. Therefore in less polar solvents the reduction in K_{ZW} is compensated for by an increase in k_r.

In Class I phosphoramidates the value of K_{ZW} is much larger than unity in aqueous solutions, therefore the transfer to a less polar solvent has a less marked effect than in Class III. As in the case of Class III phosphoramidates there is also an increase in the value of k_r. However, as the observed rate constant for Class I monoanion hydrolysis is purely dependent on k_r an increase in k_{obs} is also noted.

2.20 The structure of the transition state for monoanion hydrolysis

The increase in k_r in less polar media arose because the transition state features charge dispersal or neutralisation relative to the ground state. In a polar media a transition state with a reduced charge density would be less favoured owing to a reduction in electrostatic interactions between the solvent and ground state reagents. The opposite would be true in non-polar solvents, making the transition state more favourable and increasing the values of the observed rate constant.

As previously stated, the rate of hydrolysis of Class I phosphoramidates is very sensitive to the pK_a of the conjugate acid of the leaving amine, indicating a transition state in which the bond between the nitrogen and phosphorus atoms is substantially broken. The absence of a deuterium kinetic isotope effect (see Table:2.3) in phosphoramidate hydrolysis and observation of hydrolysis of phosphorylimidazolium and phosphoryl–*N*–methylimidazolium ions indicates that proton transfer to the amine is complete before degradation takes place and does not feature in the rate determining step.

Benkovic and Sampson postulated two mechanisms for the hydrolysis, both of which featured transition states where cleavage of the P–N bond to form an amine leaving–group is well advanced. The first of these mechanisms features the unimolecular dissociation of the zwitterionic species into a metaphosphate intermediate, as in the hydrolysis of orthophosphate esters (Fig:2.18 and 2.23). The second reaction scheme operates *via* a bimolecular displacement of the amine by a nucleophilic species (typically a solvent molecule).

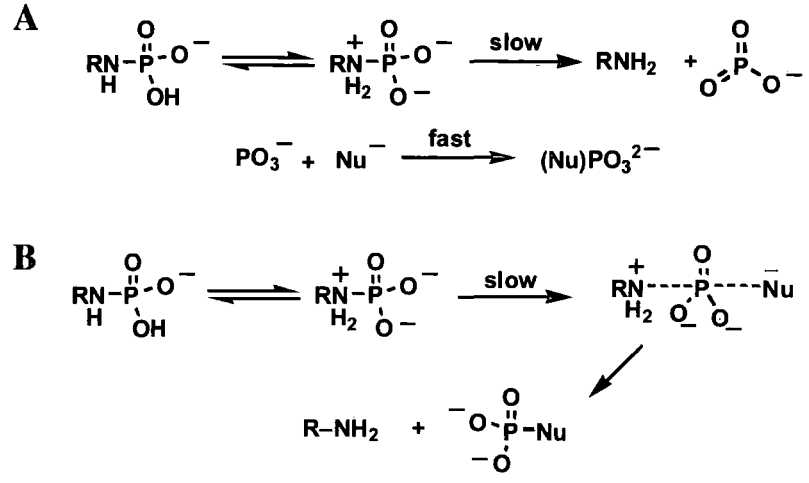


Fig:2.23 Two postulated mechanisms of phosphoramidate zwitterion hydrolysis. **A** unimolecular rate–determining step with metaphosphate formation, **B** mechanism with a bimolecular rate determining step–showing preference towards more nucleophilic phosphoryl group acceptors[50].

If the unimolecular degradation mechanism were taking place the high reactivity of the metaphosphate species would ensure the subsequent reaction with a solvent molecule would show no selectivity for differing solvents based on their respective nucleophilicities. Instead the ratio of phosphorus containing products, inorganic phosphate upon reaction with water, and phosphate ester through reaction with the more nucleophilic alcohols, would approximate the mole ratio of water and alcohol in the solvent medium. However, the experimental evidence (Table:2.3) shows a clear trend toward formation of the phosphate monoester, results which support a biomolecular reaction mechanism in which selectivity of products is based on the nucleophilicities of the attacking species.

A more exact description of the transition state was achieved through investigating the rate of reaction between various amine nucleophiles (with pK_a ranging from 2–11) with phosphorylpyridine. The rate of these reactions was correlated with a Brønsted coefficient of β =0.25, a value similar to that reported for the nucleophile–catalyzed hydrolysis of phosphoramidic acid by a series of substituted amines (β=0.2). In conjunction with this data the authors used pooled kinetic data from various sources to evaluate the Brønsted value for the complete transfer of phosphoryl anion from the donor phosphoramidate to a series of pyridines, yielding a β_{eq} =1.25 for the equilibrium process (Fig: 2.24)

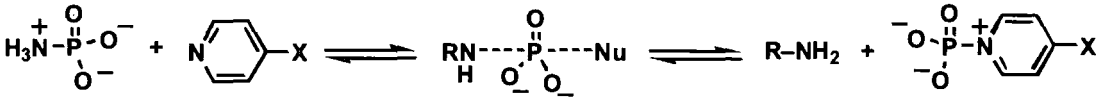


Fig:2.24 Equilibrium between phosphoramidate and phosphorylpyridinium ion., the β value of which has been estimated at 1.2.[50]

The Brønsted value is lower than that for the transfer of an acyl group to an alcoholate species (β =1.6–1.7) and only slightly higher than that for the addition of a proton (β = unity by definition). Thus within set limits the phosphoryl group can approximate the electropositive character of a proton. Within this model the fractional charge transferred to the incoming nucleophile can be estimated as 0.2/1.2 units, with a corresponding and simultaneous loss of 1.0/1.2 units of positive charge from the amine leaving–group. In keeping with these estimates Benkovic and Sampson identified the excess electron density on the anionic oxygen atoms as the driving force for the transfer of the phosphoryl group from phosphoramidate to an amine nucleophile. Assuming this model can also be applied to oxygen nucleophiles it is apparent that bond breaking between nitrogen and phosphorus (β_{lg} = 1.0) is far more advanced than the degree of bond formation between phosphorus and the incoming nucleophile (β_{nuc} = 0.2).

Therefore the description of the transition state the model predicts is loose and non–coupled *i.e.* the degree of bond breaking is not proportional to bond formation. The

model Benkovic and Sampson proposed is a bimolecular transition state that shares features with that of both the concerted mechanism and stepwise mechanistic extremes. However as the β_{nuc} and β_{lg} values show, bond formation and bond breaking are not equal in the transition state, resulting in an 'exploded' transition state wherein charge transfer is detectable, though not complete.

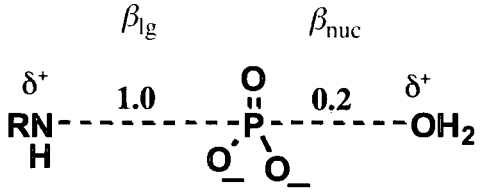


Fig:2.5 The proposed 'exploded' transition state of phosphoramidate monoanion hydrolysis[50].

The preference for phosphoramidates to react with alcohols over water can be explained in this model by theorising that the higher nucleophilicity of the alcohol allows formation of the transition state earlier along the reaction co-ordinate than is possible with water (see Section 2.22).

2.21 Corroborative evidence for a concerted mechanism

Supporting evidence for an exploded transition state during phosphoryl transfer from pyridinium phosphates (hydrolysis if water is the nucleophile) to oxygen nucleophiles, was provided by Jencks and Herschlag in their 1989 paper. A detailed analysis was laid out in which the second order rate constants for a number of phosphoryl transfer reactions between oxygen nucleophiles of varying pK_a and a range of pyridinium phosphate derivatives, also of varying pK_a.

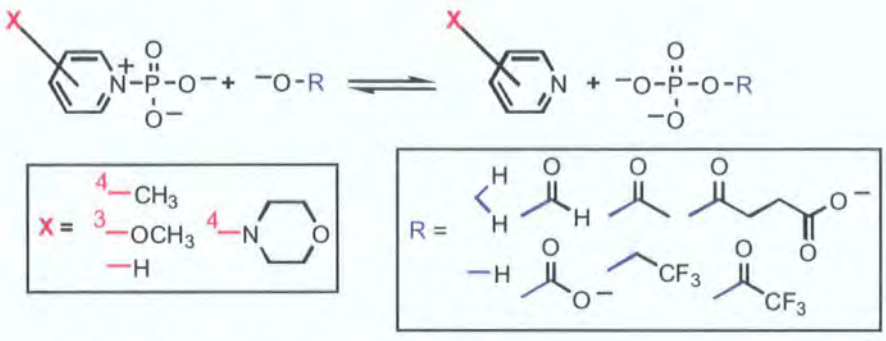


Fig:2.25 Scheme detailing the phosphoryl transfer reaction studied by Jencks and Hershlag, showing the range of pyridinium derivatives and oxygen nucleophiles used[51].

Correlations were made between the observed rate constants and the pK_{nuc}s and pK_{lg}s, to give a range of β_{nuc} and β_{lg} values. From these correlations Jencks and Herschlag were able to disprove the formation of a metaphosphate intermediate and derive evidence for the exploded transition state hypothesised by Benkovic and Sampson. They also defined an interaction coefficient linking the strength of the attacking nucleophile and the measure of P–N bond cleavage.

2.22 Postulated mechanisms

As was the case with the research carried out by Chanely and Feagson and Benkovic and Sampson, Jencks and Herschlag set out a range of possible reaction mechanisms for the hydrolysis (or phosphoryl transfer) of phosphoramidate monoanions.

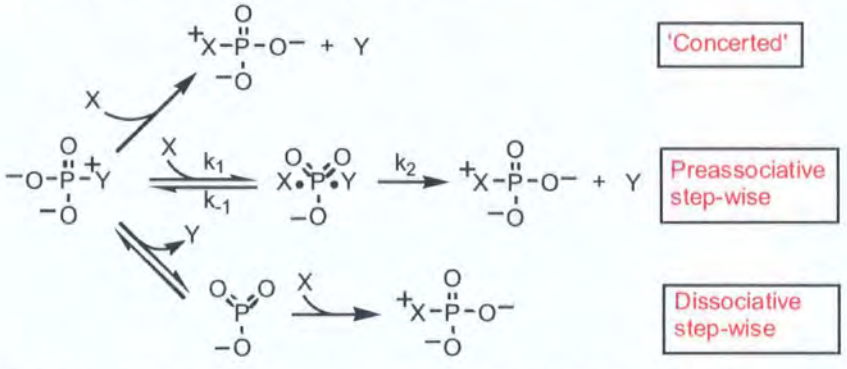


Fig:2.26 Three possible mechanistic extremes of phosphoryl transfer as proposed by Jencks and Herschlag[51].

Of the three possible routes an exploded transition state mechanism was favoured because:

1. Phosphoryl transfer reactions were second order (first order in both phosphoramidate and nucleophile), consistent with a concerted type or preassociated mechanism, but not a dissociative type.
2. If a Preassociative stepwise mechanism operates then the range variation in rate constant is limited to 10^3 by the rate constant for diffusion of an encounter pair ($\sim 10^{10}$ s^{-1}) and the frequency of bond vibration ($\sim 10^{13}$ s^{-1}). This is less than the observed variation of 10^5.
3. Jencks and Herschlag were able to derive an equation to describe the relationship between β_{nuc} and the pK_{lg}.

   $p_{xy} = \delta \beta_{lg} / \delta\, pK_{nuc} = \delta \beta_{nuc} / \delta\, pK_{lg} = \delta^2 \log k \;/\; \delta\, pK_{nuc} \delta\, pK_{lg} = 0.014$

   Where p_{xy} is the interaction coefficient. This relationship provides evidence for an exploded transition state by illustrating the link between bond formation to the nucleophile and bond breaking to the leaving group.

The interaction described in point 3 was illustrated by plotting $-\beta_{lg}$ vs (pK_{nuc} + log p/q) for a range of oxygen nucleophiles (see Fig:2.25).

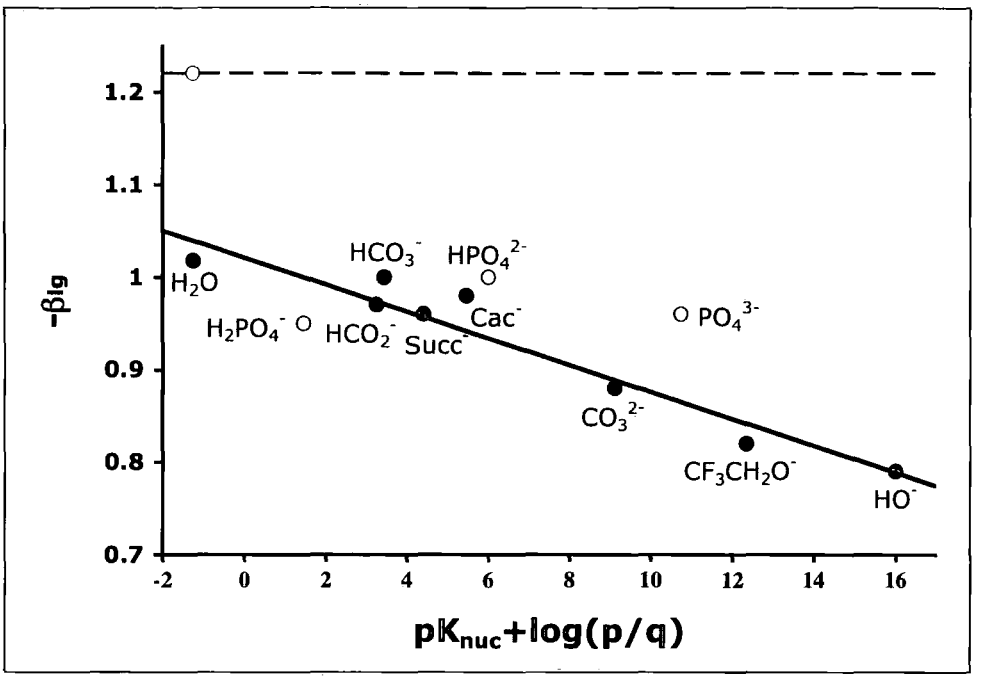


Fig:2.27 Plot of $-\beta_{lg}$ vs pK_{nuc} for reactions of phosphorylated pyridines with oxygen nucleophiles. The dashed line represents the value of $-\beta_{lg}$ expected for complete bond breaking in the transition state, the dashed circle is the value expected for solvolysis by a stepwise preassociation mechanism with metaphosphate intermediate[51].

Although Jencks and Herschlag proposed a concerted type reaction mechanism with an exploded transition state, in which β_{nuc} and β_{lg} vary according to the relationship in point 2, they also recognised conditions under which a metaphosphate intermediate could be formed.

1. In the Gas phase.
2. Stabilization by substitution of the metaphosphate oxygens with sulfur or nitrogen atoms.
3. If bond breaking and bond formation are uncoupled by a change to a weaker nucleophile (pK_{nuc} = -20) or a better leaving group (pK_{lg} = -13).

2.23 Window of opportunity

Using the data collected by Benkovic and Sampson we were able to make a preliminary calculation of the half–life of phosphoramidate hydrolysis across a range of pH levels[50]. If our intention to use a phosphoramidate masking group was to be viable we required the

half–life of hydrolysis at pH 8 to be in the order of hours rather than minutes, this would give us ample time to isolate GANP and to use it in transcription studies. A low half–life would cause problems as the phosphoramidate may not be adequately stable for use in transcription reactions *i.e.* the novel nucleotide may hydrolyse, become insoluble and precipitate out of the reaction mixture, potentially reducing the observed level of incorporation. The rate of hydrolysis at low pH is less important as within the limit of RNA stability we are free to choose the pH at which the unmasking occurs and therefore tune the rate for our own means.

Of the phosphoramidates that have had their pseudo first order k_{obs} value assessed by Benkovic and Sampson, methoxyethyl phosphoramidate bore the closest structural similarity to GANP. As shown in previous sections, the rate of phosphoramidate monoanion hydrolysis correlates to the pK_a of the conjugate acid of the parent amine. Hence, phosphoramidates which share common structural features and thereby pK_a values, can be expected to undergo hydrolysis at similar rates.

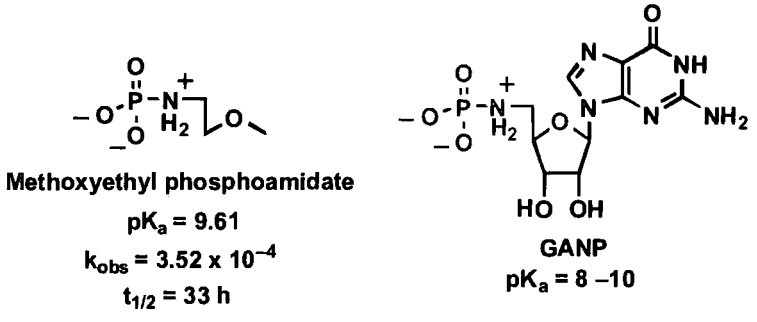


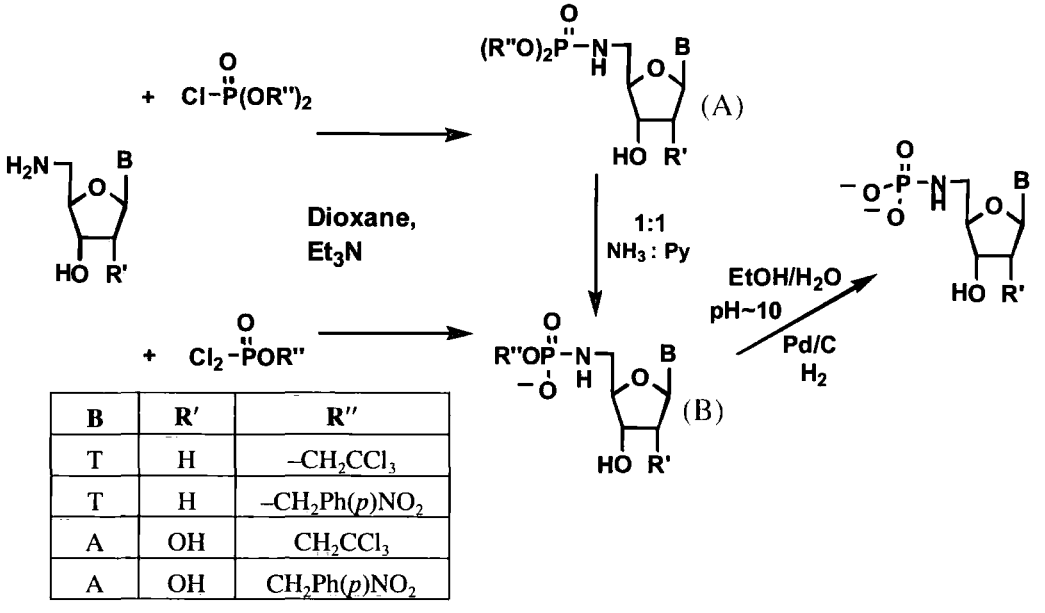
Fig: 2.28 pK_a, k_{obs} and $t_{1/2}$ values for methoxyethyl phosphoramidate. Also illustrating the similar structural elements present in GANP[50].

Using the k_{obs} value for methoxyethyl phosphoramidate (3.52×10^{-4} (min^{-1}) from Fig:2.16, Table:2.4) we can estimate the half–life of the monoanionic species hydrolysis at 33 h, indicating that the phosphoramidate group of GANP should be more than stable enough at basic pH levels to allow for use of the novel nucleotide in transcription reactions lasting 2–3 h.

Having made our hypothesis for the mode of action of GANP in the synthesis of 5′–AminoRNA and discussed the relevant aspects of phosphoramidate chemistry, the following sections will detail literature methods of phosphoramidate synthesis.

2.24 Literature methods for phosphoramidate synthesis

Through a series of articles dating from 1969 onwards Bernd Jastorff and Hans Hettler detailed methods for the synthesis of *N*-phosphorylated 5'-amino-5'-deoxynucleosides using diester-phosphochloridate phosphorylating agents[52,53].



B	R′	R″
T	H	$-CH_2CCl_3$
T	H	$-CH_2Ph(p)NO_2$
A	OH	CH_2CCl_3
A	OH	$CH_2Ph(p)NO_2$

Fig:2.29 Synthetic route to 5′-phosphoramidate nucleosides as reported by Jastorff and Hettler[52,53].

Various phosphorylating agents were used to synthesise both mono– (B) and diester (A) phosphoramidates of adenine and thymine in 50-85% yield. Hydrolysis of the first P–OR″ bond was achieved efficiently with a 1:1 mix of ammonia and pyridine at room temperature. However, the second P–OR″ is resistant to alkaline hydrolysis and, owing to the P–N bond being labile at moderate to low pH, an acid hydrolysis method was unavailable. Attempts to use Snake venom phosphodiesterase, and reductive fission with zinc both produced the 5'-amino-5'-deoxynucleoside starting material. The synthesis of the adenine and thymine phosphoramidate nucleosides was eventually achieved by

carrying out catalytic reduction of the second PO–Bn bond over palladium on carbon for the minimum amount of time in an aqueous ethanol solvent buffered to pH 10.

More difficulty was encountered in the synthesis of 5′- phosphoramidate derivatives of guanosine. Following methods similar to those used above Jastorff and Hettler were unable to isolate 5'-amino-5'-deoxyguanosine 5′-N-phosphate (GANP) but did synthesise diester phosphoramide (**X**), and 2′,3′- isopropylidene (**Y**) derivatives shown below:

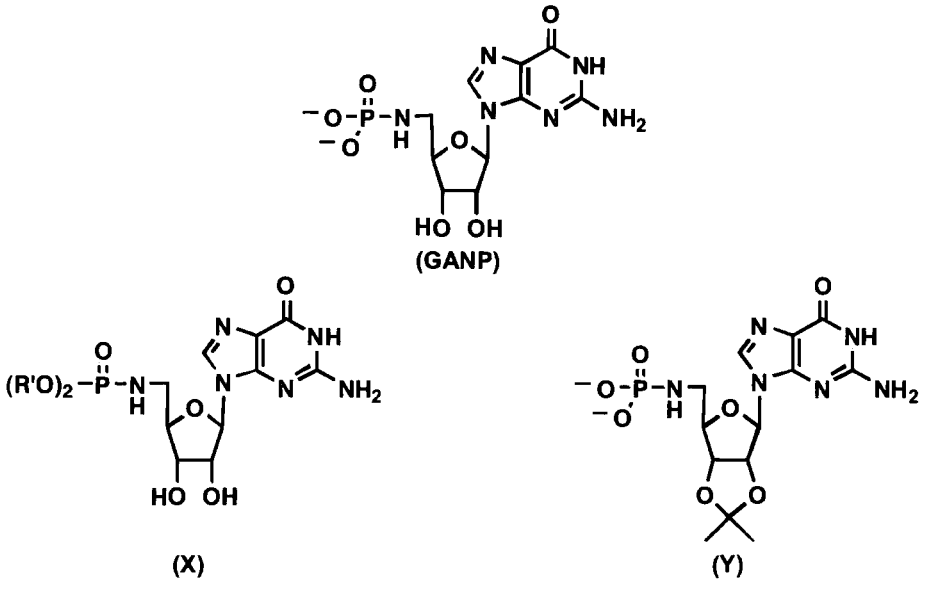


Fig:2.30 GANP and its Diester (**X**) and 2′,3′-isopropylidene (**Y**) protected derivatives[52,53].

Unfortunately neither of these novel nucleotides would be useful to our project. The diester phosphoramidate does not possess any charged groups at the 5′- position would likely possess aqueous solubility lower than that of Amino–G. The 2′,3′ – isopropylidene protected derivative does possess the necessary anionic groups, however, with the 3′ - hydroxy group blocked it is unable to initiate polymerisation reactions. Removal of the isopropylidene protecting group is usually achieved under acidic conditions, which when attempted also removed the phosphoryl masking group to yield Amino–G.

We were disinclined from attempting to adapt this procedure for our own means as:

1. Synthesis of GANP had been unsuccessfully attempted by this method in the past.
2. The multiple steps required would likely reduce yield of phosphoramidate.
3. Some of the phosphorylating agents used were uncommon and would require synthesis before use, further complicating the synthetic route.

2.25 Amino–phosphorylation approach

The strategy we devised was to first synthesise Amino–G *via* a literature method, then develop a regiospecific amino–phosphorylation reaction to give GANP. Owing to the hydrolytic instability of the phosphoramidate group we aimed to avoid the use of protecting group chemistry in the phosphorylation step. Therefore removing the necessity for further deprotection steps that had the potential to reduce our yield of GANP, and would certainly add to the complication of the synthetic route. Avoidance of protecting group chemistry makes the phosphorylation step challenging, as the amine starting material possesses a number of potentially nucleophilic groups (Fig2.28)

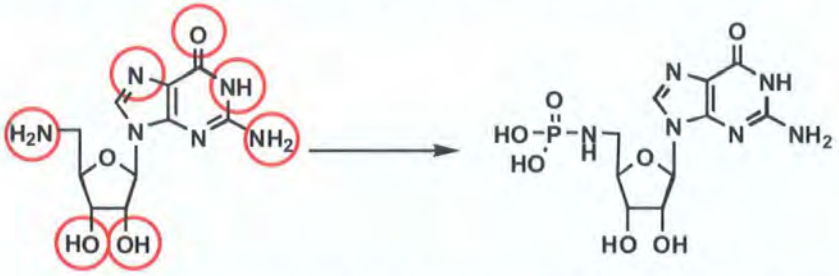


Fig:2.31 Regiospecific phosphorylation to convert Amino–G to GANP. The potentially nucleophilic groups of Amino–G are ringed

We felt this route was reasonable, as literature precedent exists for the synthesis of a close analogue of GANP, 5′–amino–5′–deoxyinosine 5′–*N*–phosphate, by a related approach. Also, Yoshikawa *et al.* have succeeded in developing a method for the regiospecific phosphorylation of the 5′–oxygen of guanosine without the use of protecting groups. Thus at the outset we appeared to have precedent for two possible routes to GANP. First I shall discuss the literature route to Amino–G, then I shall systematically work through our synthesis of the amine, and subsequent work towards devising a robust and regiospecific amino–phosphorylation reaction to yield GANP.

2.26 Synthesis of 5'-amino-5'-deoxyguanonsine (Amino–G)

The literature precedent that exists for the synthesis of Amino-G is that of D.K. Dean, which gives the final product in good yield (39%) over three steps without the use of complicated purification steps such as chromatography[46,54]. In keeping with our objectives the route is not synthetically challenging, hopefully giving any resulting methods broader appeal. The initial step is the substitution of the 5'–hydroxyl of guanosine for an iodine atom, which is subsequently substituted for an azide group in the second step. The 5'-azide then undergoes a Staudinger reduction to give Amino–G.

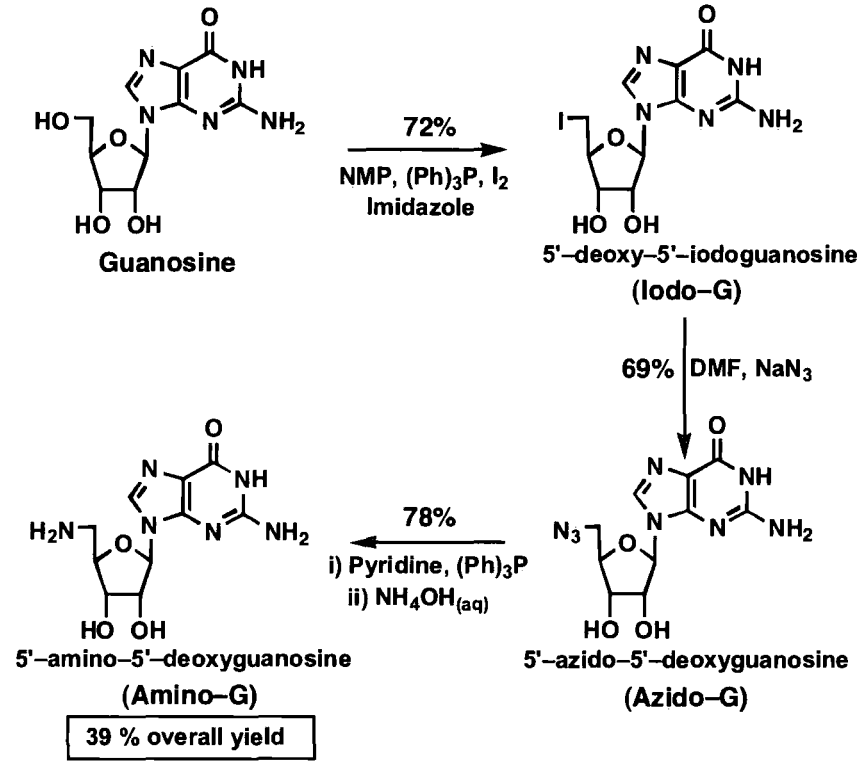


Fig:2.32 Reaction scheme for synthesis of 5'-amino-5'-deoxyguanosine (Amino–G). Yields shown are those quoted in the literature[46,54].

The following sections will discuss the chemistry behind each transformation from guanosine to GANP.

2.27 Synthesis of 5'–deoxy–5'–iodoguanosine (Iodo–G)

The work of Dean cites an article by McGee *et al.* for the synthesis of the 5'-iodo-derivative of guanosine (Iodo–G)[47]. Substitution of guanosine's 5'-hydroxyl for an iodine

atom is necessary as hydroxyl anion is a poor leaving group and therefore unsuitable for an azide group substitution.

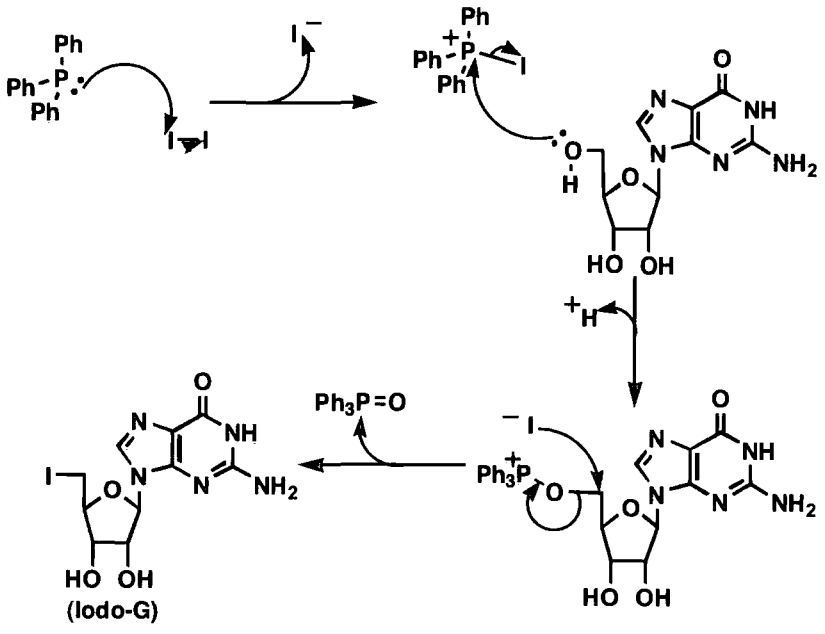


Fig:2.33 Proposed reaction mechanism for substitution of the 5'–hyroxyl of guanosinewith iodide, operating *via* a triphenylphosphonium intermediate.

2.28 Purification and characterisation of Iodo–G

Through our initial attempts to synthesise Iodo–G we discovered the reaction was sensitive to the presence of H_2O. Therefore all subsequent reactions were carried out using dried solvents, reagents and glassware. Precipitation of Iodo–G from the crude reaction mixture was achieved by addition of water and dichloromethane (DCM), with precipitate forming at the phase boundary. Crude product was collected by filtration at the pump as a pale yellow solid, ^{1}H NMR spectra of which showed the presence of an impurity as a singlet peak at 7.6 ppm, tentatively assigned to residual triphenylphosphine. Removal of the impurity was achieved by refluxing the crude product in a small volume of DCM for approximately 30 mins. Filtration at the pump yielded Iodo–G as an off-white solid. Cooling the water-DCM precipitation step to 5 °C for 48 hours to maximise precipitation, raised initially low yields to an average of 65%.

^{1}H, ^{13}C NMR, mass spectrometry, IR and melting point analysis were used to confirm that the synthesis was successful. The most convincing evidence aside from the appearance of

a correlating signal in the mass spectra came from the ^{13}C NMR spectra. The substitution of the 5'-hydroxyl of guanosine for iodine resulted in a shift in the 5'–methylene carbon signal from 61.4 ppm to 8.00 ppm. The shift of the 5'–methylene protons in the ^{1}H NMR was also helpful, but proved more difficult to observe the signal lay under a residual water peak in the DMSO-d_6 solvent.

2.29 Synthesis of 5'-azido-5'-deoxyguanosine (Azido–G)

As with Iodo–G the synthesis of Azido–G followed Dean's procedure[46]. The reaction follows a nucleophilic substitution mechanism, with azide anion acting as the nucleophile with attack being directed by the electronegative iodine atom.

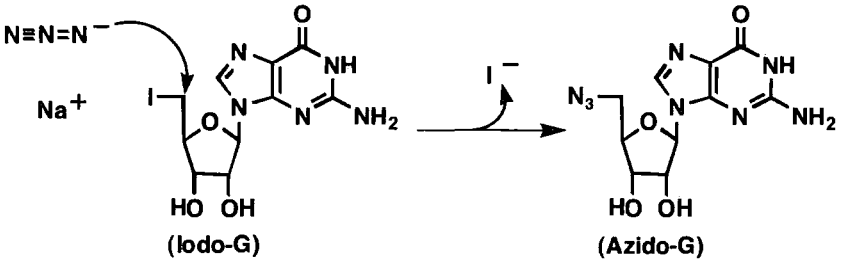


Fig:2.34 Proposd mechanism of Azido–G formation, nucleophilic attack at the 5' carbon by an azide anion.

The reaction is carried out in dimethylformamide, which as a dipolar aprotic solvent causes the azide to be present as a naked anion with enhanced nucleophilicity.

2.30 Purification and characterisation of Azido–G

The first attempts at synthesis were successful; a high yield of product was isolated as an off-white powder with no impurities present. Removal of excess inorganic azide and DMF was achieved *via* a series of solvent washes. A pure sample of Azido–G as a pale brown powder was collected and dried over phosphorus pentoxide to give average yields of 76%. Spectroscopic analysis of the product revealed the material was pure and therefore no further purification was required. The high yield and purity of the azide synthesised was such that no optimization of the reaction was deemed necessary.

A characteristic shift of the 5'–methylene peak in the ^{13}C NMR spectra from 8.00 ppm to 51.7 ppm provided strong evidence that the synthesis was a success. In conjunction with ^{1}H NMR, mass spectrometry, infra–red (a characteristic N_3 stretch was observed) and melting point data we identified the reaction product as Azido–G.

2.31 Synthesis of 5'-amino-5'-deoxyguanosine (Amino–G)

The final step towards the synthesis of Amino–G was the reduction of Azido–G's azide group to give a terminal amine. The reaction takes place in pyridine and is hypothesised to occur *via* the mechanism detailed below:

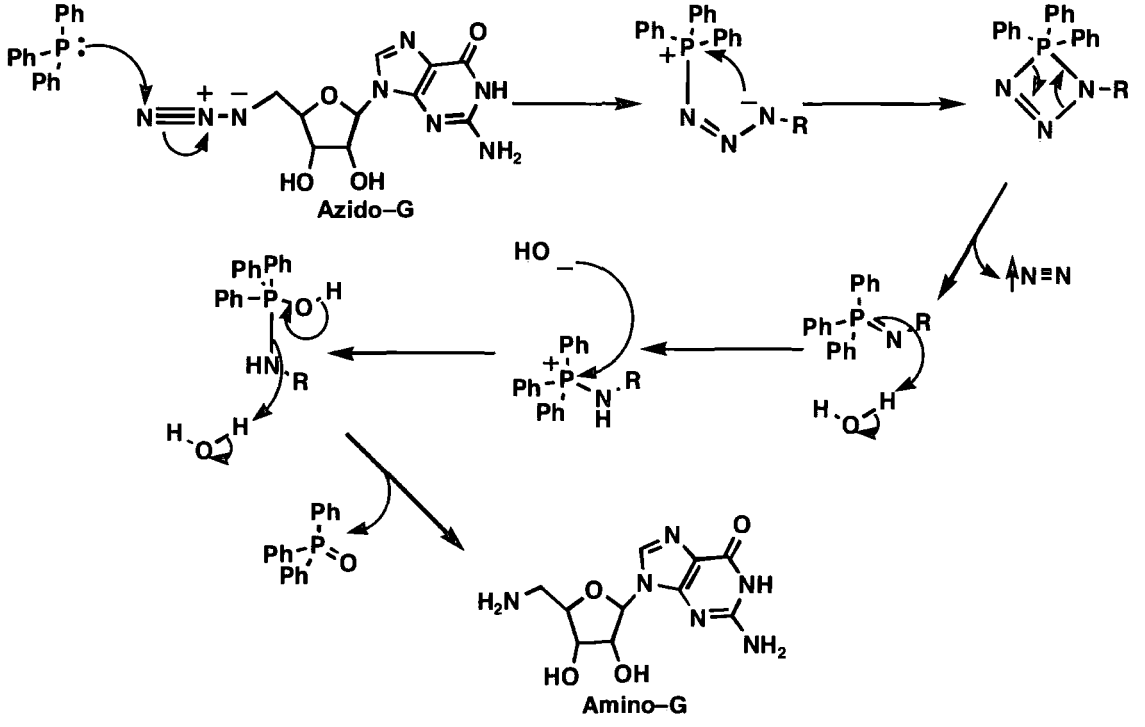


Fig:2.35 Hypothesised mechanism of reaction for the synthesis of Amino–G. Formation of the reactive triphenylphosphine adduct is proposed to take place *via* a four membered transition state.

The driving force for the reaction is the formation of the very stable phosphorus – oxygen double bond (P=O = 544 kJ mol^{-1}).

2.32 Purification of Amino–G

The work-up procedure used by Dean (2002) used a number of solvent washes to remove residual reagents/ by-products[46]. This method was successful, if rather longwinded. Through the course of our research we observed that Amino-G is partially soluble in water, especially at higher temperatures. This observation led us to successfully recrystallise the amine from water, removing the need for all but one of the solvent wash steps from the original work up. Product was isolated as small yellow crystals that were positively identified as 5'–amino–5'–deoxyguanosine by ^{13}C, ^{1}H NMR, melting point analysis, elemental analysis, and mass spectrometry. On average Amino–G was obtained in 77% yield.

With the amine starting material in hand we were now equipped to progress our research to develop a method of phosphorylation, and thus achieve GANP synthesis.

2.33 Synthesis of 5′-amino–5′–deoxyinosine 5′–N–phosphate

Hampton *et al.* reported the synthesis of 5′–amino–5′–deoxyinosine 5′–*N*–phosphate by a chemospecific phosphorylation of 5′–amino–5′–deoxyinosine[47]. They performed the reaction in dry acetonitrile solvent at 0 – 5 °C, using phosphorus oxychloride as the phosphorylating agent, pyridine was also present to act as a nucleophilic catalyst and/or base. The authors described a two–step purification process, the first stage of which was the precipitation of crude product as a barium salt. The crude barium salt was then re–suspended in the presence of K^+ ion–exchange resin, which the authors proposed had the effect of slowly dissolving the phosphoramidate as a K^+ salt. The final stage of purification was paper electrophoresis carried out in a formate buffer quoted to be at pH 8. Upon reviewing the work–up procedure, and taking into consideration the hydrolytic nature of phosphoramidate groups (Section 2.4), we find it highly suspect that after undertaking a high temperature (~100 °C) purification method such as paper electrophoresis the authors were able to isolate much, if any, phosphorylated material. Thus we used the phosphorylation method as reported, but adopted alternate strategies for purification.

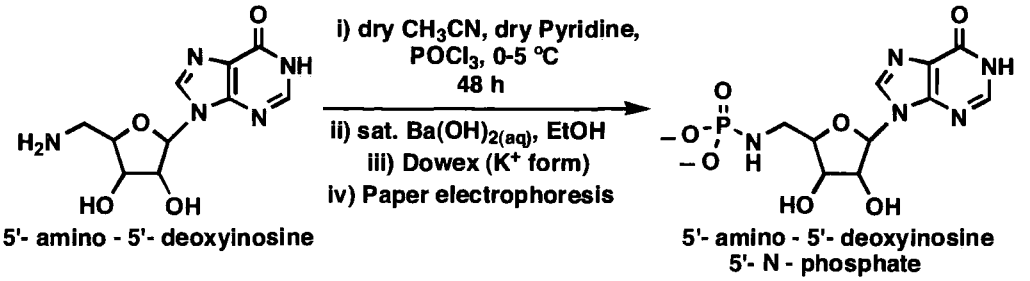


Fig:2.36 Conditions for regiospecific phosphorylation of 5′-amino-5′-deoxyinosine as reported by Hampton *et al*[47].

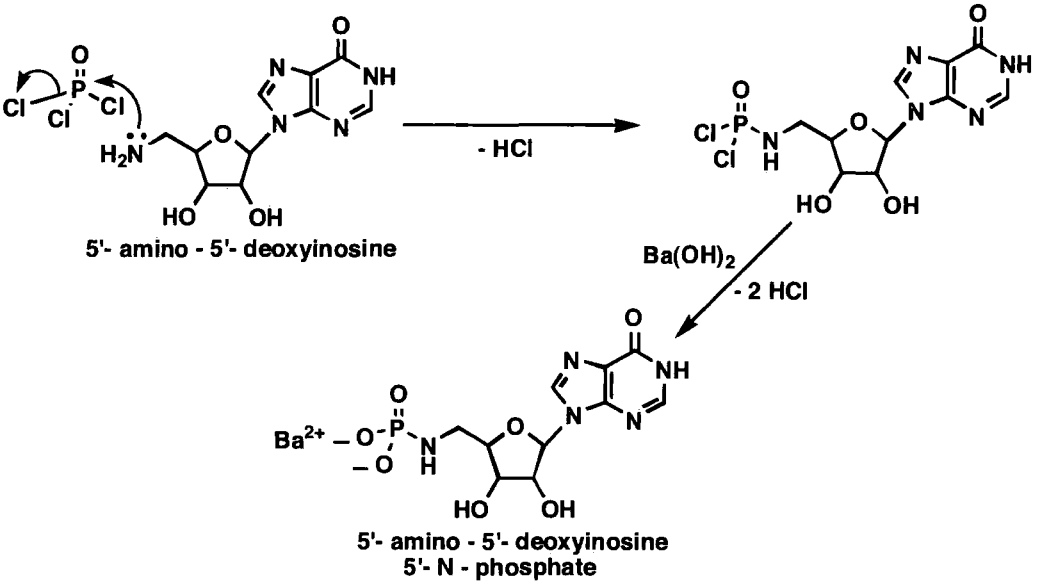


Fig:2.37 Proposed reaction scheme for the synthesis and purification of 5′-amino-5′-deoxyinosine 5′-N-phosphate.

Advantages:

- Very close analogue to the transformation we aim to achieve.
- No use of protecting groups.
- All reagents commercially available.

Disadvantages:

- Protracted work-up.
- No yield given in literature method.
- Little spectroscopic/ physical data that the synthesis was a success.
- Serious questions exist about the validity of the work-up procedure.

Upon weighing up the pros and cons of the procedure, we decided to attempt modifying this reaction for our own means. The overwhelming reason was the obvious similarity between this reaction and the phosphorylation we hoped to achieve. The main

disadvantage was the length of the work-up, including both ion exchange and electrophoresis steps. As previously mentioned we aimed to keep the synthesis as simple as possible by avoiding lengthy purification techniques. However we felt these two steps could be condensed into one through the use of anion-exchange chromatography. As previously discussed (Section 2.23) we anticipated the $t_{1/2}$ for phosphoramidate hydrolysis to be on the order of hours in a solution of neutral pH, a time period we believed to be more than adequate to allow elution from a column of ion exchange media. Through the use of chromatography rather than electrophoresis we also gained the significant advantage of a lower operational temperature. The rate of phosphoramidate hydrolysis is highly sensitive to temperature, and while electrophoresis inherently operates above ambient temperature it is possible to cool chromatographic columns, thereby further extending the half–life of any product as it passes through the column.

2.34 First approach towards GANP synthesis

All reagents and solvents were dried before use (see Section 9.1) to avoid side reactions between water and the phosphorylating agent.

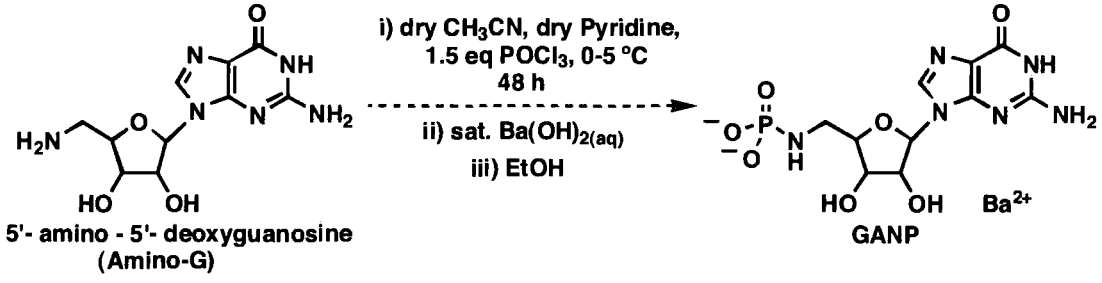


Fig:2.38 Proposed phosphorylation reaction as modified from Hampton conditions.

Following incubation the resulting thick suspension was added to a solution of saturated barium hydroxide. Presumably this step was intended to hydrolyse the remaining P–Cl bonds, while maintaining a high pH to stabilise any product. The solution was left for 30 mins, during which time a precipitate (suspected to be the barium salt of inorganic phosphate) formed. The solid was removed by centrifugation before the next step of nucleotide purification from the solution was attempted. As barium salts of nucleotides are known to be insoluble in ethanol–water mixtures we attempted to precipitate our

desired product through the addition of two volumes of ethanol to the remaining solution. Upon ethanol addition and cooling the solution to –20 °C a white precipitate was formed.

Once isolated and dried under vacuum, the solid proved to be essentially insoluble in aqueous solution, however we attempted to dissolve in D_2O to collect NMR spectra, though no clear signals were present in the ^{13}C, ^{1}H, or ^{31}P spectra. A number of possibilities exist that could explain these observations:

1. Synthesis was unsuccessful resulting in no product being present in the reaction mixture.
2. Reaction was a success and GANP was precipitated as a barium salt upon quenching the reaction mixture in the sat.$Ba(OH)_2$ solution, and discarded with the other insoluble barium salts.
3. GANP was precipitated from the aqueous ethanol solution, but remained insoluble and therefore not present in the samples used to collect NMR spectra.

As we were uncertain of the relative solubility of our product in comparison to the barium salt of inorganic phosphate (formed through reaction of residual phosphorus oxychloride with the aqueous quench solution) and could not be sure when our product would be precipitated, we chose to find an alternative to the barium precipitation method. Our inability to analyze all the products of the precipitation by any convenient method, owing to the innate insolubility of barium salts, was key in making this decision.

2.35 Alternative to Barium precipitation

We required reagents that would hydrolyse both the remaining phosphorus oxychloride and the dichloro–phosphoramidate intermediate, while maintaining a high pH and our ability to analyse the resulting products. We decided to quench the reaction mixture in a 150 mM solution of sodium carbonate. We expected this to meet all our criteria, as all products formed should remain soluble as sodium salts and the pH would remain high. After quenching, separation of our product from by-products and impurities would be achieved by anion-exchange chromatography. The isolation of the phosphoramidate from the eluent solution could then be achieved through selective precipitation methods.

As expected, upon quenching no precipitate was formed and a reduction in the solution's pH was observed. The resulting clear yellow solution was loaded on to a Sephadex DEAE A25 anion exchange column. Sephadex Diethylamine ethylene (DEAE) A25 is a chromatography media that can be used to separate a mixture of compounds according to the charge they carry. The media possesses a large number of ammonium groups that partake in electrostatic interactions with anionic compounds that have been loaded on to the column. The bound compounds are eluted from the column by increasing the salt concentration of the eluent over time. Materials possessing a larger negative charge are bound more tightly to the column and therefore require a higher concentration of salt to displace them. Uncharged or cationic species are subjected to no impedance by the A25 media and therefore pass directly through the column. A flow-UV detection system was used to track the elution of products and reagents.

Separation was attempted utilizing a triethylammonium bicarbonate (TEAB) gradient running from 0.1 – 0.5 M at pH 7.5. Data regarding the rate of phosphoramidate hydrolysis provided in Benkovic and Sampson's 1971 paper allowed us to estimate the half-life of an aryl phosphoramidate to be ~ 35 h at pH 8 (Section 2.23)[50]. If the hydrolysis of GANP took place over a similar time scale we should have ample time to carry out the necessary chromatography and isolate our product without a major reduction in yield.

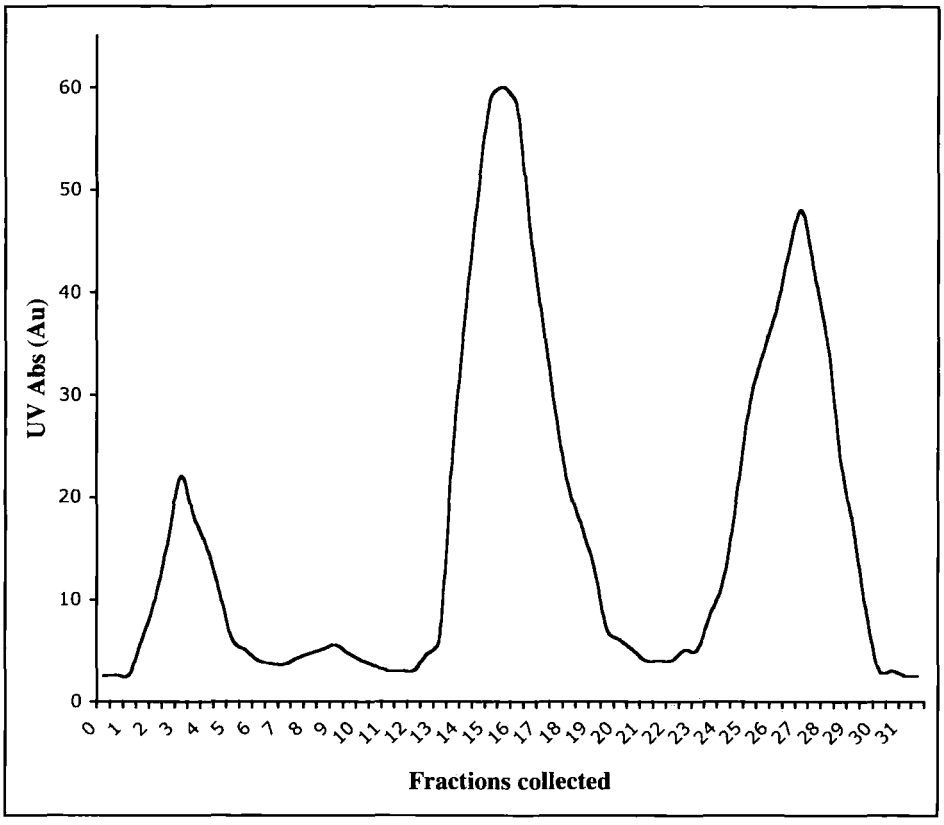


Fig:2.39 Chromatogram showing elution of UV active material from a DEAE A25 Sephadex column.

The resulting chromatogram indicated a number of UV-active substances had eluted from the column. The fractions corresponding to the UV-active substances were lyophilized to give white solids containing a large proportion of salt, analysis of which by mass and NMR spectroscopy techniques provided no evidence that synthesis had been successful.

In an attempt to improve separation we altered variables in the system. This included a reduction in salt concentration during quenching, using an ammonium bicarbonate eluent running over a low gradient, and loading less crude product onto the column. The use of ammonium bicarbonate rather than the triethylamine bicarbonate buffer allowed us to analyse ^{1}H NMR spectra without the presence of signals from ethyl groups of TEAB. Changes to the gradient and sample loading resulted in a marginally clearer chromatogram, allowing us to identify individually recurring peaks over a number of

experiments. However the strength and resolution of signals in the ^1H NMR spectra were still too poor to afford the identification of individual substances. In an effort to remove the impurities from the eluted samples we decided to subject them to a second chromatographic column.

Removal of small molecule impurities, such as salt and inorganic phosphate, was attempted by gel filtration chromatography using a Sephadex G10 matrix with water as the eluent. Sephadex G10 is a porous media that retards the elution of small molecules while larger compounds are unaffected. Our aim was to lyophilize the collected fractions from a DEAE A25 column, which even though not fully resolved should have achieved some separation of the anionic product from the neutral starting material. Lyophilized fractions were redissolved in a small quantity of water and loaded on to the G10 column.

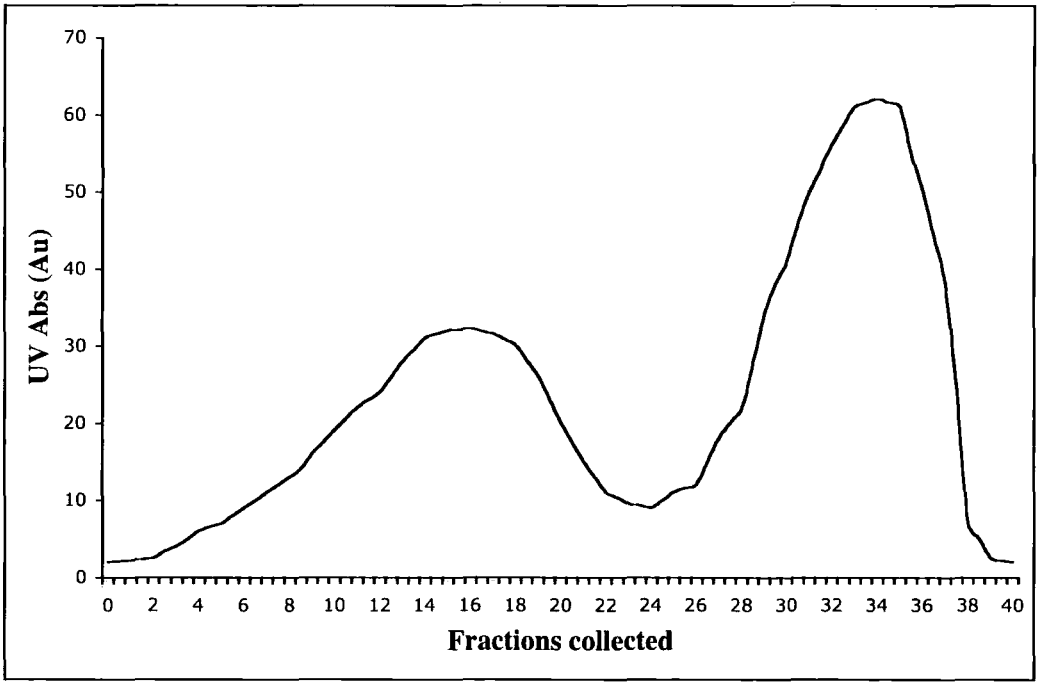


Fig:2.40 Plot of eluted fractions vs UV activity for G10 gel filtration column

The presence of multiple UV signals in the chromatogram indicated the gel filtration was achieving further separation of the UV active substances. These signals were again

poorly resolved, a problem we attempted to correct by altering the amount of sample loaded on the column and by adjusting the eluent flow-rate. Although improvement was made, resolution of peaks was still poor. Samples recovered from the G10 column were lyophilized and showed a visible reduction in salt content compared to those collected from the A25 column. However owing to very low mass recovery neither NMR nor mass spectra contained signals of sufficient intensity to indicate the presence of GANP. A number of low intensity signals were observed in the ^{1}H NMR spectrum, but we could not positively identify any of the substances present. No signals were observed in the ^{31}P NMR spectra other than inorganic phosphate.

As time taken to work-up and purify the reaction mixture was becoming overly extensive and fruitless, we decided a change was needed. We could see no fault in the logic for the work-up method, other than our initial assumptions regarding the $t_{1/2}$ of phosphoramidate hydrolysis were incorrect. This was unlikely given the number of independent publications quoting similar rates of reaction, therefore it was decided that an alternative phosphorylation procedure be used instead[48-51].

2.36 Yoshikawa method of regiospecific nucleic acid phosphorylation

We abandoned the phosphorylation method by Hampton *et al* as we had observed no indication over many experiments that the reaction was working. Our lack of faith in their results and the long reaction time also played a part in our decision to move in a new direction. We opted to utilize a phosphorylation method developed by Yoshikawa *et al* to phosphorylate the 5′– oxygen of nucleosides without the use of protecting group chemistry[55,56]. The selectivity of this reaction is excellent, favouring phosphorylation at the 5′– oxygen over the only slightly less nucleophilic 2′ and 3′ positions. As we require the reaction to be as selective as possible to minimise purification steps, and thereby maximise yield, adapting this procedure for our own means seemed to be a logical step. As with the previously attempted phosphorylation reaction, phosphorus oxychloride was used as the phosphorylating reagent. However, the solvent used was a trialkyphosphate, which the authors postulate reduces the activity of $POCl_3$ by forming an ionised structure.

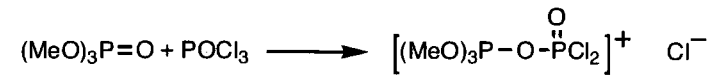


Fig:2.41 Trimethylphosphate reduces the activity of Phosphorus oxychloride[55].

Standard reaction time for the Yoshikawa method was much shorter than the Hampton procedure at only 7h. Also, a good deal more data was available for the use of the reaction with a variety of substrates under differing conditions. Although the method has not been employed for the phosphorylation of amino–nucleosides we would expect high levels of conversion owing to the high nucleophilicity of amines in comparison to hydroxyls. The supposed reaction mechanism for the Yoshikawa method is almost identical to that used for the Hampton reaction, with the exact nature of the phosphorylating reagent the only discriminating feature. In a slight variation to the method used by Yoshikawa *et al.* we added an equivalent of pyridine to our reaction mixture to act as a base. As a work-up the reaction mixture was quenched in triethylamine solution and loaded on to a Sephadex A25 column. Products of the reaction were then eluted using an ammonium bicarbonate gradient.

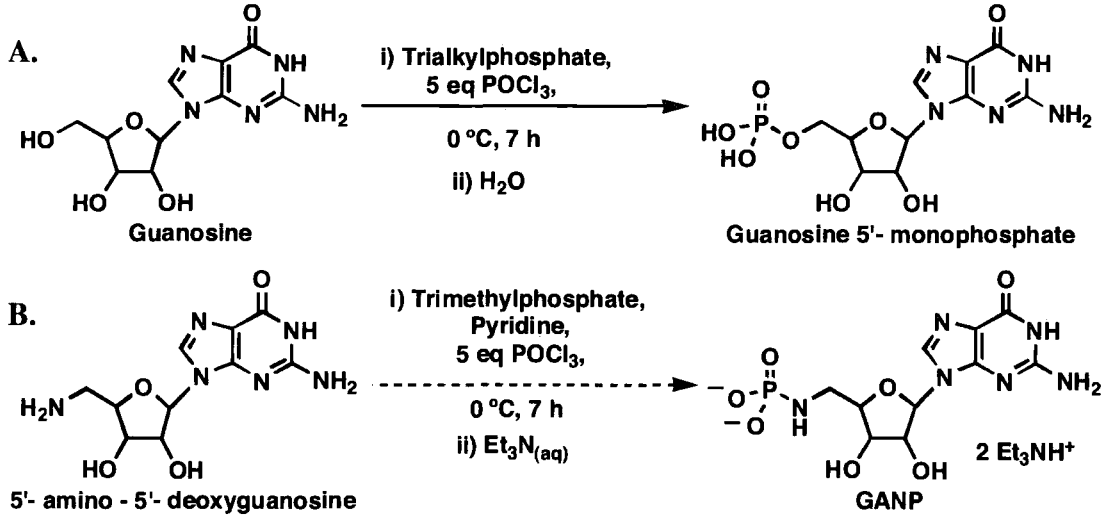


Fig:2.42 **A.** Conditions for the highly regiospecific nucleoside phosphorylation as reported by Yoshikawa, **B.** Yoshikawa method adapted for GANP synthesis[55,56].

2.37 First attempt at Yoshikawa based approach

Our initial attempts using the Yoshikawa method seemed promising. Upon ion-exchange chromatography of the quenched reaction mixture a well resolved chromatogram was

produced, the peak of highest intensity was the final substance eluted. Considering the dianionic nature of GANP this result indicated the method had potential. However NMR and mass spectrometry indicated that the signal was not GANP.

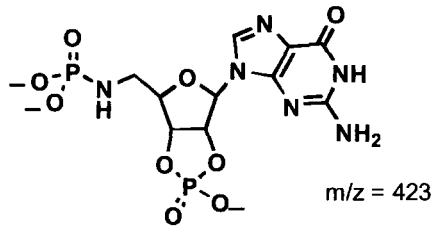


Fig:2.43 Hypothesised major product of Yoshikawa reaction,
structure corroborated by mass spectrometry and NMR

Electrospray mass spectrometry of the major product gave a spectrum in which two peaks dominated. One of lesser intensity at 423 m/z is tentatively attributed to the structure given above, while a signal of higher intensity at 343 m/z fits the same substance after loss of PO_3H from the 5′–phosphoramidate. Signals in the ^{31}P NMR appear to confirm the presence of this, or a similar product, however no conclusive assignment has been made. Other peaks in the chromatogram showed the presence of a guanosine derivative as low intensity signals in ^{1}H NMR spectra. However ^{31}P NMR of these samples gave spectra where the only signals were indicative of trimethylphosphate and inorganic phosphate.

2.38 Optimisation of reaction

The production of a diphosphorylated nucleotide suggests the use of excess phosphorylating reagent. To reduce the occurrence of this impurity we chose to reduce the number of equivalents and/or reactivity of phosphorus oxychloride in the reaction mixture. Although a reduction in the amount of $POCl_3$ used in the reaction mixture resulted in a decrease in the amount of impurity observed, as determined by ^{31}P NMR spectroscopy and mass spectrometry, the only other signals present were those assigned to inorganic phosphate and the trimethylphosphate solvent.

2.39 Reduction of the reactivity of phosphorus oxychloride

Articles by Yoshikawa state an increase in selectivity of phosphorylation reaction can be gained by the introduction of alcohols or water in controlled amounts to the reaction mixture. The effect of these additions is to reduce the activity of the phosphorus oxychloride by substitution of one or more chlorine groups to form a phosphate monoester dichloridate.

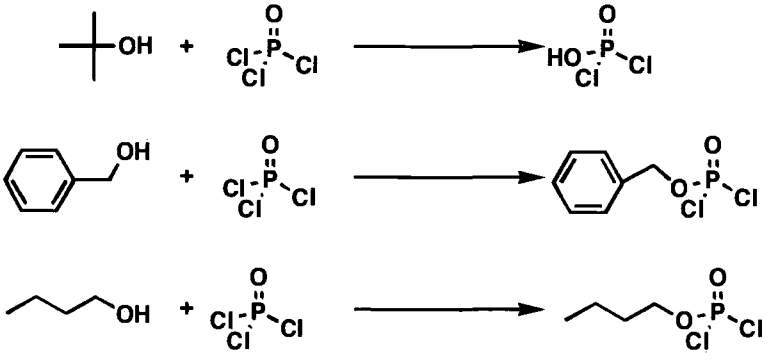


Fig:2.44 Use of t-Butanol, Benzyl Alcohol, n-Butanol to reduce activity of Phosphorus oxychloride[55].

Additions of benzyl alcohol and t-butanol to the reaction mixture did not enable us to isolate any compounds likely to be ester analogues of GANP. However the addition of n-butanol to the reaction mixture resulted in the appearance of a novel peak in the chromatogram.

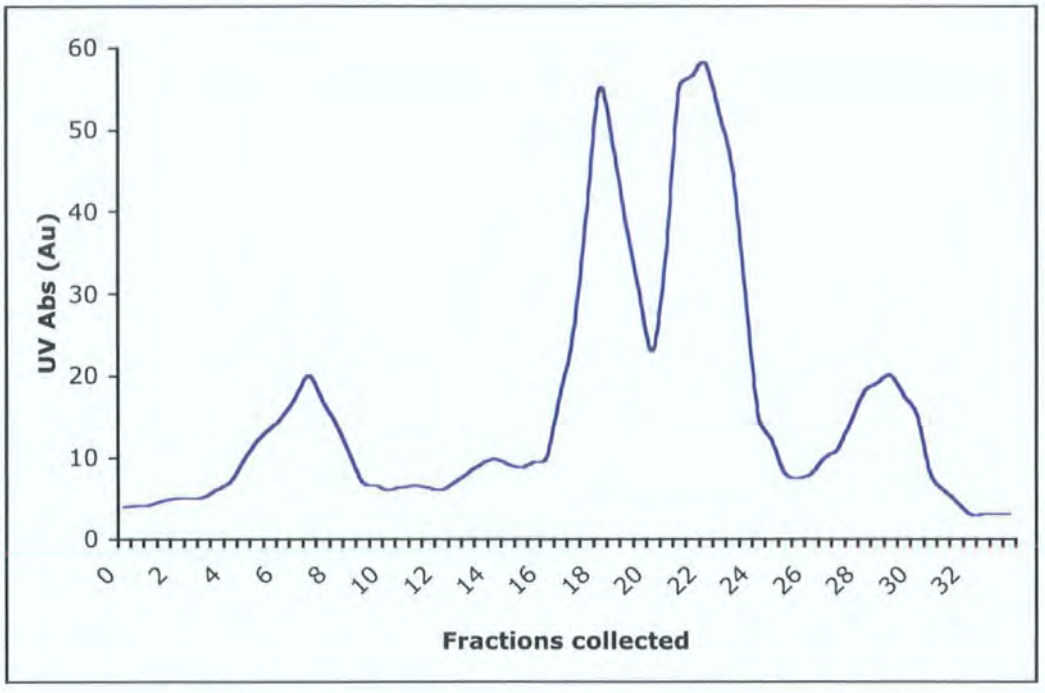


Fig:2.45 Chromatogram showing UV absorbance *vs* fractions collected using DEAE A25 Sephadex. An extra peak can be seen *c.f.* Fig:2.36

Judging by the elution positions for neutral substances such as pyridine and for more anionic substances such as the suspected diphosphate impurity, the substance corresponding to the peak carried a low negative charge. ^{31}P NMR spectrum indicated the presence of a phosphoramidate with characteristic signal at 10.1 ppm[57]. Mass spectrometry corroborated our suspicion that the product was in fact the n-butyl phosphoramidate ester.

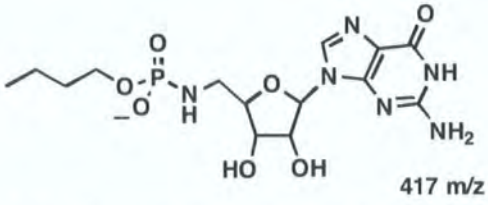


Fig:2.46 n-butyl phosphoramidate guanosine derivative synthesised through addition of n-butanol to Yoshikawa reaction mixture. 5′ group not anionic enough to serve our requirements, but is encouraging in that the synthesis does indicate regiospecific phosphorylation is possible.

Although not the required product the isolation of this compound indicated that the regioselective phosphorylation of 5'-amino-5'-deoxyguanosine was possible. While

investigating phosphorylation reactions based on Yoshikawas conditions we were aware the possibilities for adapting the procedure for our needs were running out. We widened our search through literature procedures to encompass not only nucleoside phosphorylations but also all amine phosphorylations in general that resulted in a phosphoramidate product. Amongst others this search yielded the protocol by Drueckhammer detailed below[58]. Our first attempt at the synthesis of GANP using a reaction based on this method was so successful that we halted research into Yoshikawa type methods at this point. Although we had the method using a trialkylphosphate solvent as a back-up procedure, it was never necessary to continue our research in this area owing to the success with which the aqueous reaction scheme was employed.

2.40 Aqueous Phosphorylation reaction

Upon gaining poor results from procedures sourced in nucleic acid chemistry *i.e.* Hampton and Yoshikawa methods, we widened our search to encompass all examples of amino-phosphorylation. Consequently we found a promising method in an article by Drueckhammer in which a chemospecific amino-phosphorylation reaction was carried out in aqueous solution[58]. The reaction was developed for the phosphorylation of a terminal amine in alkaline conditions through the gradual addition of phosphorus oxychloride diluted in dry THF. The chemospecificity of the phosphorylation is attained through taking advantage of the amines exceptional nucleophilicity by gradually addition of the dilute $POCl_3$. By utilizing a 'semi–batch' method reaction between $POCl_3$ and the more nucleophilic amine is favoured, while side reactions occurring between $^-OH/H_2O$ and $POCl_3$ are reduced.

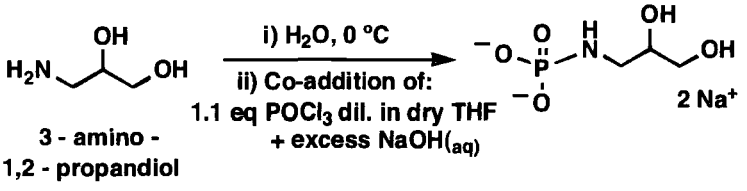


Fig:2.47 Chemospecific amino-phosphorylation of 3-amino-1,2-propandiol in aqueous conditions[58].

As the reaction occurs under aqueous conditions the dichlorophosphoramide, formed in the first step of the reaction hydrolyses to form the phosphoramidate very quickly. The

short lifetime of the dichlorophosphoramide reduces the possibility of a cyclic phosphodiester being produced through an intramolecular reaction with the hydroxyl group at the 2 position (see Fig:2.48) *i.e.* $k_{inter} >> k_{intra}$.

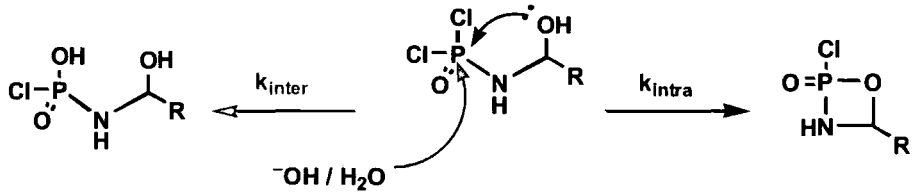


Fig:2.48 Two routes of reaction for a dichlorophosphoramide, one of which is bimolecular with a rate constant of k_{inter} the second route being unimolecular with the rate constant k_{intra}. If k_{inter} is larger than k_{intra} we would expect more phosphoramidate than the cyclic phosphorester to be produced.

The acidic species produced during the reaction were neutralized by the addition of 6M $NaOH_{(aq)}$, maintaining the high pH conditions of the reaction mixture. As we will go into more detail below, the amount of hydroxide present in the reaction is key in determining the yield of phosphoramidate produced. Therefore the pH of the reaction was closely monitored to determine when NaOH additions were necessary.

In our own application of this method we expect the only by–products of the reaction to be NaCl, NaOH and inorganic phosphate. We believed that GANP could be purified *via* its precipitation as a sodium salt from an ethanol:water mixture. The potential ease of gaining pure product factored in with the simple, yet elegant, reaction scheme attracted us to investigate further the phosphorylation of Amino–G by this method. However, we believed our starting material to have limited solubility in aqueous solution. Obviously this makes the literature method, as it stands, inapplicable for our needs. Hence an alteration was made to the protocol through the introduction of a 2:1 H_2O:DMSO co-solvent. The Amino–G starting material had already displayed high solubility in DMSO, thus we expected to be able to dissolve the amine in this good solvent, before diluting into the poor aqueous solvent. Upon dissolution a pale yellow solution was formed with no visible precipitation. Upon cooling the solution noticeably thickened, most likely owing to the high melting point for DMSO. No change was observed in the reaction mixture

during the majority of the $POCl_3$ addition, however as the final 10 % was added an amount of white solid was formed.

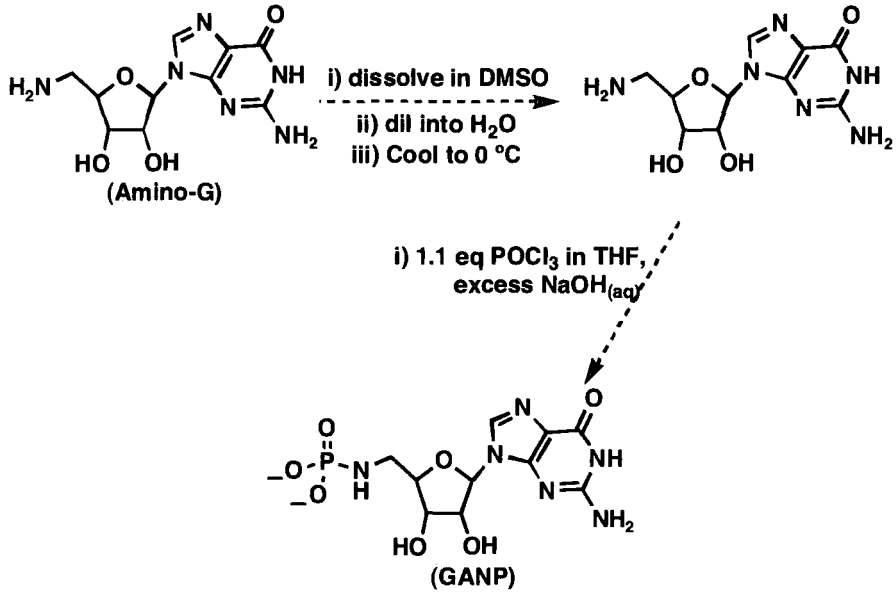


Fig:2.49 Proposed reaction scheme. DMSO is present to facilitate the solvation of Amino–G into aqueous solution.

2.43 First attempt at aqueous method

Once addition of the phosphorylating agent had been completed the solvents were removed from the reaction mixture under reduced pressure, leaving a white solid. This had not been possible when using the Yoshikawa phosphorylation method as the boiling point of the trimethylphosphate solvent used was too high. By isolating the crude reaction mixture in a hydrolytically stable form at high pH (checked by pH meter) we were able to analyse it using a range of analytical techniques. The first of which was ^{31}P NMR spectroscopy, the resulting spectrum exhibited two signals in approx. 1:2 intensity ratio. The smallest of these peaks was at ~ 8.5 ppm, while the second was at ~ 2.3 ppm. The peak at 8.5 ppm is in the correct range for a phosphoramidate group, more evidence for identification of the signal was acquired by collecting the ^{31}P NMR spectra showing ^{1}H coupling[57,59].

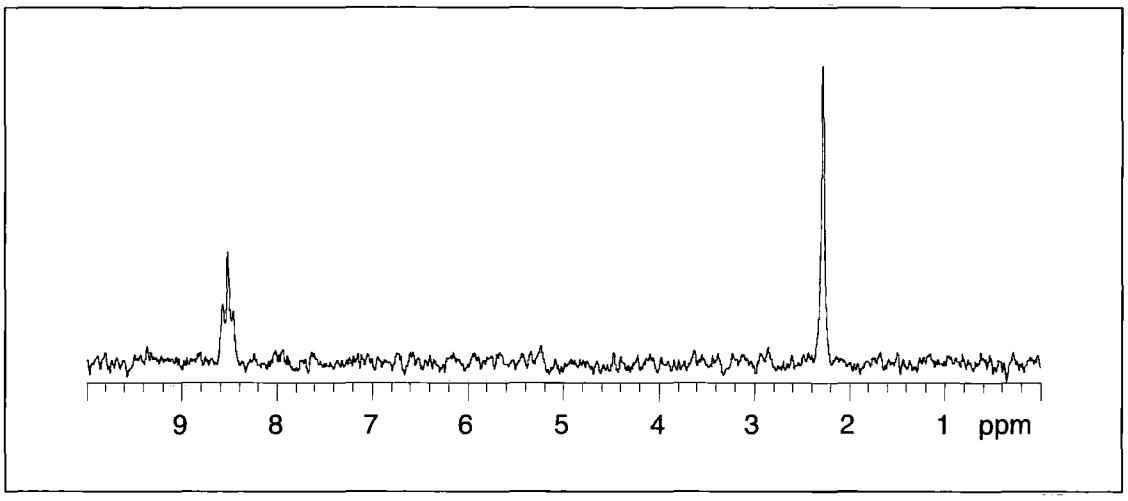


Fig:2.50 ^{31}P NMR spectrum of the crude reaction mixture showing ^{1}H coupling.

The ^{1}H coupled ^{31}P NMR spectra showed two signals; coupling was observed in the signal at 8.5 ppm, splitting it to a triplet. The peak at lower shift showed no splitting, indicating that no ^{1}H atoms were attached within 1–3 bonds of the phosphorus nucleus that was responsible for this signal. By analysing the possible phosphorylation sites on the Amino-G starting material we can determine which would give rise to a triplet signal in the NMR spectra. Splitting of signals owing to nearby ^{13}C or ^{15}N isotopes is rarely observed in ^{31}P NMR spectra[59]. This is because the low natural abundance of these isotopes makes any coupling present too weak to resolve. Those protons attached to the nitrogen atom do not couple to the phosphorus atom owing to their high rate of exchange. Hence we can be confident that the triplet signal observed in the ^{31}P NMR spectra originates from a phosphorus group coupling to the 5′ methylene of GANP.

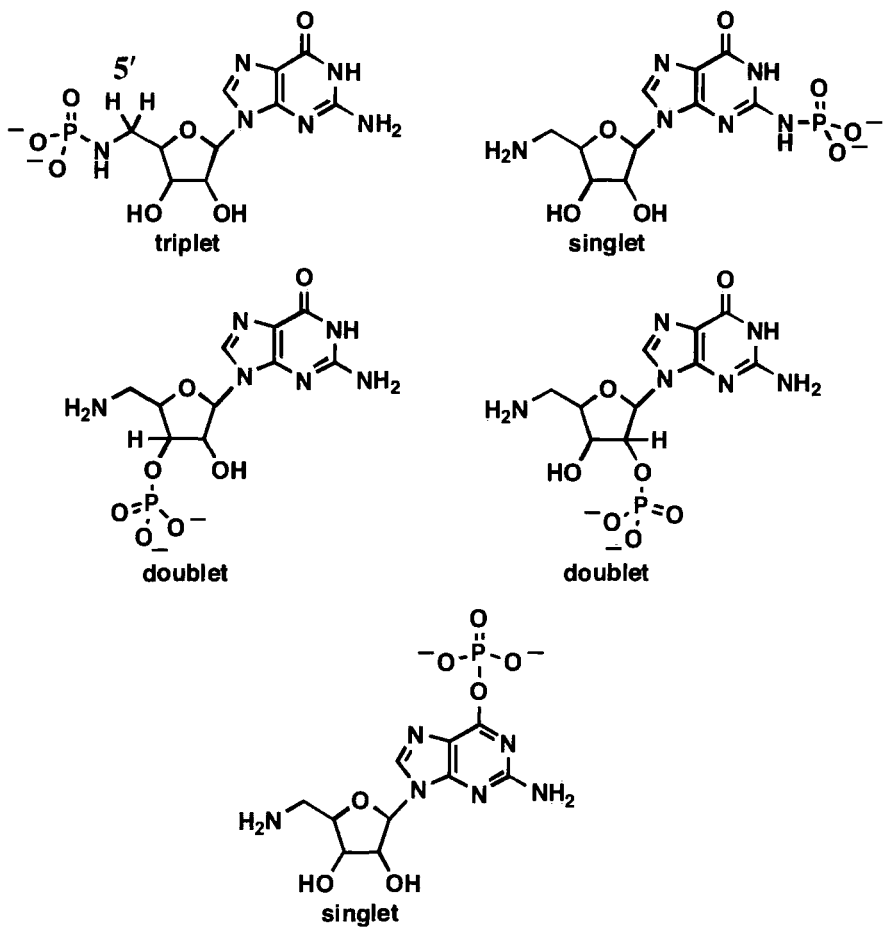


Fig:2.51 Potential structures of phosphorylation reaction products. The most nucleophilic group of Amino–G is the 5′–amine, making GANP the most favoured product of the phosphorylation reaction.

As the diagram above shows, only the phosphoramidate of GANP is likely to result in a triplet signal in the ^{31}P NMR spectra. Therefore the ^{31}P NMR spectrum of the crude reaction mixture confirms that the selective amino–phosphorylation reaction had been a success. The presence of a signal at 361 m/z in the LC – mass spectrum provided corroborating evidence of GANP's presence. Unfortunately we were unable to identify any product by ^{1}H NMR as a large peak from residual water in the sample dominated the spectra, ^{13}C NMR was excluded because at this stage we lacked the quantity of sample necessary to collect a strong spectrum.

The low chemical shift and lack of proton coupling observed in the second signal in the ^{31}P NMR spectrum together with knowledge of the possible by–products of the reaction

led us to suspect the presence of inorganic phosphate[59]. Our suspicions were confirmed through making two significant observations.

2.44 Addition of inorganic phosphate to crude product

The first of these was a 'spiking' experiment wherein a sample of crude reaction mixture was analyzed by ^{31}P NMR spectroscopy collected, once before and once after the addition of a known quantity of inorganic phosphate. The pre–addition spectra showed the presence of two signals as expected, one triplet peak assigned to GANP, and a singlet peak at lower chemical shift. If the unknown signal originated from inorganic phosphate, the addition of inorganic phosphate to the sample would increase the intensity of this signal, while that of GANP would remain unchanged. If inorganic phosphate were not the source of the unknown signal, an additional peak would be observed in the second ^{31}P NMR spectrum. As expected an increase in the lower shift signal was observed after addition was made, confirming our suspicions. The assignment was still not conclusive however; as it remains possible that the unknown material could merely produce a signal co-incident to that of inorganic phosphate. For this reason a second experiment was undertaken to corroborate our observations.

2.45 Tracking GANP hydrolysis by ^{31}P NMR spectroscopy

It is known that the hydrolysis of phosphoramidate groups produces terminal amines and inorganic phosphate. As these two materials produce signals of differing chemical shifts in ^{31}P NMR spectra it is therefore possible to assess the relative amounts of the two phosphorus containing compounds in a given sample. We hypothesised that given these two factors it would be possible to incubate a sample of the crude reaction product (containing both GANP, and the unknown material) under conditions favouring hydrolysis (neutral pH, ambient temperature), then by collecting a sequence of ^{31}P NMR spectra, over time we could track the conversion of GANP to inorganic phosphate. If the unknown phosphorus containing material were inorganic phosphate we would expect to see an increase in signal at lower shift as the GANP signal decreased in intensity. Again if the unknown material were not inorganic phosphate an additional peak would be

observed. Though this experiment was measuring a similar change to the 'spiking' experiment, we felt it was more conclusive as it was a more passive process, *i.e.* we were not adding anything to the sample, merely tracking an inherent change. In corroboration with our previous results and suspicions an increase in the signal of lower chemical shift was observed as the GANP signal was seen to decrease. This strongly indicated that hydrolysis of the phosphoramidate was taking place and that inorganic phosphate was present in the crude reaction product (see Sections 2.3–23 for a detailed discussion of phosphoramidate hydrolysis).

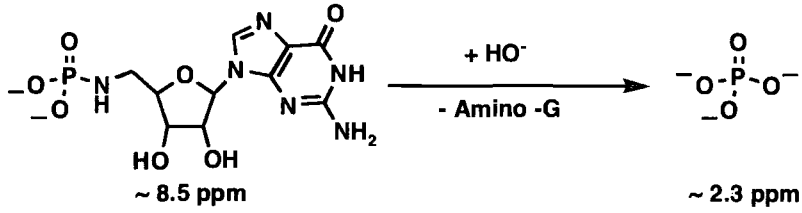


Fig:2.52 Hydrolysis of the phosphoramidate group of GANP to form inorganic phosphate can be observed *via* ^{31}P NMR. Accurate integral measurements can be made owing to the significant change in chemical shift as illustrated above.

2.46 Source of inorganic phosphate

The presence of inorganic phosphate in the crude product indicates that we may not have been phosphorylating all of our amine starting material. Therefore with an aim of optimising the phosphorylation, and a view to both increasing yield and eliminating the need to remove a potentially troublesome impurity, we examined potential sources of inorganic phosphate production:

- Impure phosphorus oxychloride; this was possible though the reagent was distilled prior to use, making it an unlikely occurrence.
- The THF in which the phosphorus oxychloride was diluted could have been wet, producing the impurity before amino–phosphorylation was possible; the use of rigorously dried solvent and oven dried glassware during the reaction would eliminate as much of this side reaction as possible.
- The hydrolysis of GANP during the reaction; extremely unlikely given the high pH, low reaction time, and low temperature reaction conditions (see Section 2.4)

- Side reactions occurring between H_2O/ ^-OH and $POCl_3$; we believed this to be the most significant source of the impurity as both nucleophiles were present in the reaction mixture.

2.47 DMSO:water optimisation study

We hypothesised that as the amine starting material was the best nucleophile in the reaction, the large amount of phosphate produced had its source in a lack of availability of the amine. Although ^-OH is also a good nucleophile, and was present in excess of the amine, we reasoned that as most $POCl_3$ added to the reaction was converted to GANP, the amine was a significantly more reactive species in this reaction. Low availability of starting material could be owing to the amine group being protonated and therefore non-nucleophilic, or the starting material having precipitated as the reaction progressed. The pK_a of terminal ammonium ions is usually in the range 8 – 9 and as the pH of the reaction was monitored and maintained above 10, the nucleophilicity of the amine was assured. The solubility of the starting material however was less clear-cut. As previously mentioned the amine is sparingly soluble in neutral aqueous solutions but soluble in DMSO. However upon cooling the reaction mixture the solubility of the starting material in the co-solvent system could become compromised. Therefore we designed and ran experiments to investigate the effect that proportions of DMSO: H_2O had on product distribution, hoping to find an optimum where solubility reached a maximum.

Phosphorylation reactions were carried out under several sets of conditions where temperature, quantity of starting material, timing of $POCl_3$ additions and amount of H_2O/ NaOH in the reaction mixture were all varied. An excess amount of NaOH was used and added to each reaction mixture before addition of $POCl_3$ began, thereby maintaining high pH conditions without varying the amount of NaOH present in otherwise comparable reaction mixtures. Although this results in the use of a large excess of hydroxide with respect to the amine and phosphorylating agent it did provide a level of consistency between runs that would be hard to attain through manual co-addition of base and

phosphorus oxychloride. The amount of DMSO used was varied from 0 – 50 % (v/v) (using amounts over 50 % resulted in the reaction mixture freezing upon cooling to 0 °C).

% DMSO (v/v)	DMSO (μL)	H_2O (μL)	% GANP yield
0	0	400	69
10	40	360	61
20	80	320	48
30	120	280	62
40	160	240	55
50	200	200	48

Table:2.4 Variation in GANP yield as % DMSO is altered, as determined by ^{31}P NMR.

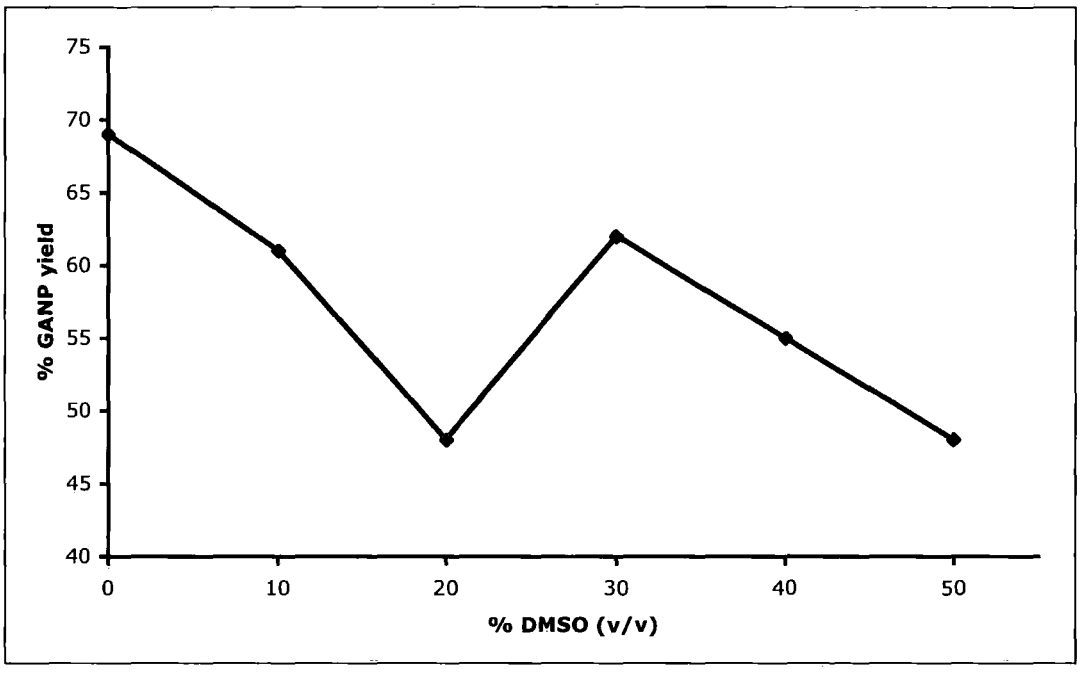


Fig:2.53 Plot of GANP yield as determined by integration of ^{31}P NMR vs % DMSO (v/v) used in the reaction mixture. No linear relationship can be inferred from the data.

2.48 Conclusions from DMSO study

The investigation did not allow us to draw a simple linear relationship between the percentage of DMSO used, and the yield of GANP. However, a very important

observation was made regarding the phosphorylation using 0 % DMSO. We had expected the starting material to be completely insoluble under these conditions and therefore the resulting yield of GANP to be the lowest observed. Contrary to our expectations this experiment gave the highest yield of GANP, Amino-G being completely soluble in the fully aqueous solution.

As suspected the starting material was initially insoluble in the aqueous solvent, but on addition of the excess NaOH, however, complete dissolution of the solid occurred. Continuing the reaction as normal no irregular observations were made *i.e.* no precipitation of solids, no significant change in pH. This implied that the removal of DMSO from the reaction mixture had no adverse effect on the reaction.

What this experiment clearly demonstrated to us is that:

1. DMSO is not necessary in the reaction mixture as the starting material has a pH dependent solubility in aqueous solution.
2. A factor other than solubility of the amine could be controlling product distribution.

2.51 The solubility of Amino-G

As a first step towards understanding the factors that contribute to the reaction yield we considered why Amino-G becomes soluble in aqueous solution upon addition of base. The most logical explanation is that a deprotonation occurs, creating a charged group on the compound and thereby making it more soluble in polar solvents such as water. By assessing the pK_a's of groups in Amino–G we can form an opinion of which are likely to deprotonated in the presence of excess hydroxide.

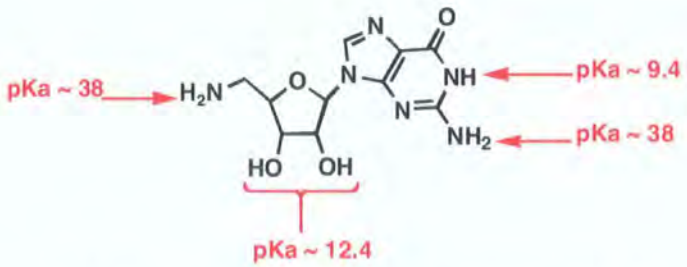


Fig:2.54 pK_a values of some Amino-G functional groups[32].

It is apparent that with a pK_a value of 9.4 it is the nitrogen at position 5 on the guanine base (N^5) that is most likely to deprotonate. A low pK_a value can be justified at this position through the delocalisation of the resulting negative charge.

2.52 Reassessment of key factors in aqueous reaction

In our original hypothesis the yield of GANP was determined by solubility of the starting material, which we believed to be a factor of the amount of DMSO present in the reaction mixture. This was shown to be incorrect by our DMSO investigation though we did not rule out the solubility of Amino-G as an important factor in product distribution. If all the amine was available to react with the $POCl_3$, yet not all the amine was reacting, it follows that a competing reaction was taking place between the $POCl_3$ and another reagent to form inorganic phosphate, thereby limiting the amount of $POCl_3$ available to produce the phosphoramidate. Both H_2O and ^-OH are present in the reaction mixture and could react with $POCl_3$ to form inorganic phosphate.

We hypothesised that through carrying out an Amino-G phosphorylation reaction in which water was excluded we could assess the degree to which a side reaction between $POCl_3$ and water was taking place. The exclusion of water was achieved through undertaking the reaction in a 100% DMSO solvent (at room temperature) and using the same amount of NaOH (added as a solid on this occasion) used in previous experiments. We expected that if water were taking part in a competing reaction a decrease would be seen in the amount of inorganic phosphate produced and GANP yield would increase (making the tenuous assumption that reactivity is equal in aqueous and DMSO solvents). Phosphorus NMR spectroscopy of the crude reaction product showed GANP was

produced in a yield (56%), not dissimilar to results gained for the mixed solvent reactions. This result indicated that any competition between water and Amino-G was negligible, and that the low GANP yield must have an alternate source.

The only species remaining in the reaction mixture capable of reacting to form inorganic phosphate is hydroxide anion. Although less hydroxide was present in the reaction mixture than water, owing to its full negative charge it is a better nucleophile. To investigate the effect hydroxide has on reaction yield we designed and ran a series of experiments varying the concentration of sodium hydroxide used while keeping all other variables constant.

2.53 Hydroxide concentration investigation

There is a minimum amount of hydroxide necessary for the phosphorylation reaction to be successful, through a series of experiments we aim to find this optimum quantity. The most successful reaction will be that which produces the highest percentage yield of GANP, assessed *via* ^{31}P NMR spectroscopy. We believed the optimum concentration of hydroxide could be determined by a number of factors key to the reactions success:

1. Dissolution of starting material,
2. All HCl produced must be neutralised.
3. Final reaction mixture at high pH (9+) where GANP is relatively stable.

We hypothesised that a minimum of 4 equivalents of hydroxide with respect to the amount of starting material used (3 to neutralise HCl produced, 1 to maintain pH and dissolve Amino–G) would be required for the reaction to be produce GANP. Therefore the number of equivalents of base studied was in the range 4–6.5. As with the previous study all other variables (amount and concentration of starting material, $POCl_3$ addition rate, and reaction temperature) were kept constant throughout the experiments.

Eqs NaOH	% GANP	% Inorganic Phosphate	% Other Impurities
4	47	41	12
4.5	88	9	3
5	91	9	Trace
5.5	85	9	6
6	80	11	9
6.5	55	25	20

Table:2.5 Change in product distribution as Eq NaOH is altered.

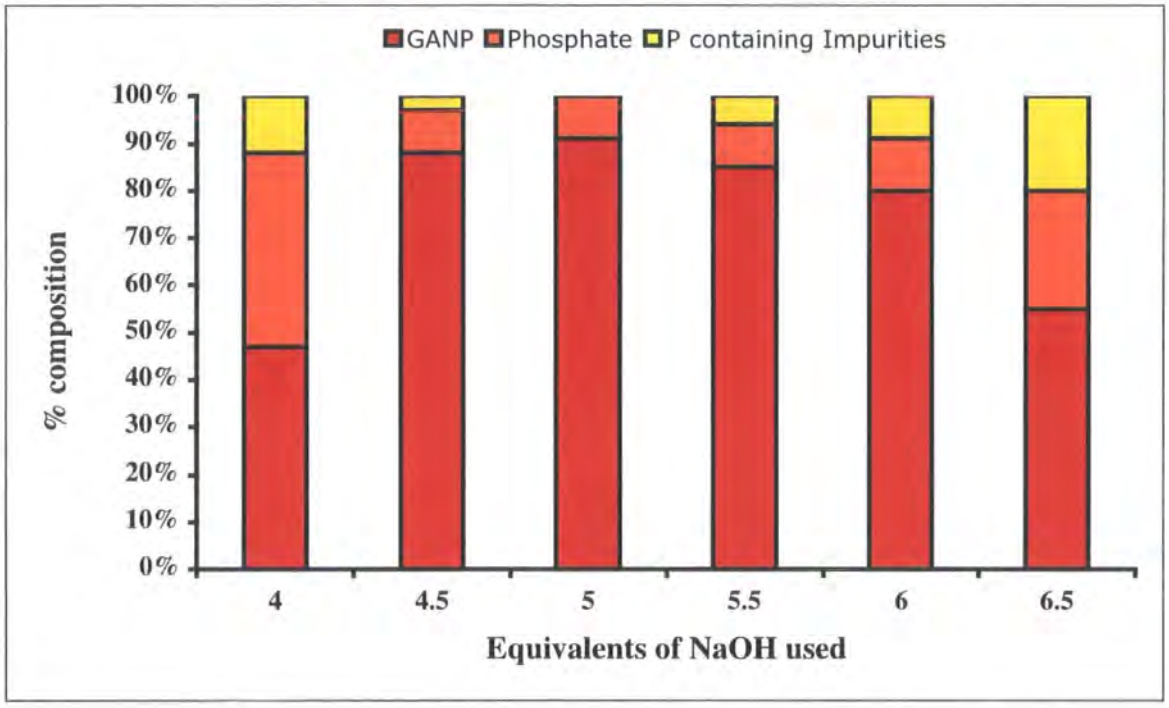


Fig:2.55 Percentage yield of phosphorylation products vs equivalencies of NaOH used. Yields determined though integration of ^{31}P NMR signals.

As can be seen in the results given above we determined that by using 5 equivalents of base the yield of GANP is maximised while minimising the formation of phosphate and other phosphorus containing impurities. These other impurities had been present in previous experiments, but at such low intensities that they remained uncharacterised. The identification of an optimum point, where GANP yield is almost double that observed previously, indicates that our hypothesis was correct and that hydroxide concentration makes an important contribution to the phosphorylation reaction's selectivity.

2.54 Identification of reaction products

The use of lower concentrations of hydroxide resulted in the precipitation of solid throughout the reaction and a low final pH, which contributed to a poor GANP yield. Higher equivalencies gave no precipitate and maintained high pH, however the increased amount of hydroxide available to take part in side reactions resulted in lower yields of GANP and higher yields of phosphate and other impurities. In the experiment performed within the range 4.5 – 5.5 equivalents of NaOH the GANP yield was essentially invariant, however there was variation in the yield of inorganic phosphate and other phosphorus containing impurities. Impurity signals were observed in the ^{1}H coupled ^{31}P NMR spectrum at 4.5 and 3.8 ppm, both signals exhibited proton coupling to form doublets. Looking at possible phosphorylation sites (as discussed above, Section 2.25) on Amino-G indicates two sites, which if phosphorylated would give doublet signals:

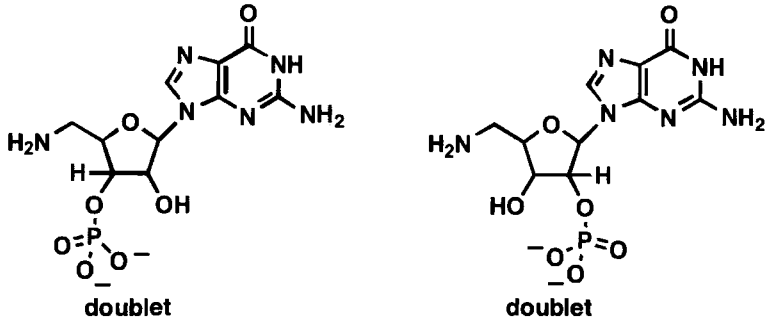


Fig:2.56 Structures of hypothesised phosphoester impurities. The chemical shift and coupling of signals in the ^{31}P NMR spectra provide corroborating evidence for the presence of these by products.

We can propose a reaction scheme whereby nucleophilic attack by the 2′,3′ hydroxyls of Amino-G at $POCl_3$ could form the phosphate monoesters. Although we were unable to formally characterise the impurity signals owing to their low abundance in the reaction mixture, the chemical shifts in the ^{31}P NMR spectra are corroborated by literature values, and the splitting pattern is consistent with phosphorylation at the 2′,3′ hydroxyls. Evaluation of the pK_a values for nucleophilic groups in Amino-G also aids in identifying those positions where reaction may occur. The pK_a of the 2′,3′ hydroxyls is 12.4, hence

as the pH of the solution is increased towards this value, a higher proportion of the more nucleophilic oxyanion will be formed. We would therefore expect to see an increase in the yield of the 2′ and 3′-phosphoesters as pH increases, within the limit of competition with hydroxide anion itself. This hypothesis is corroborated by our hydroxide study as shown in Table:2.5.

Hence we can be confident in our assignment of the impurities as the 2′ and 3′-phosphate monoester products of side reactions.

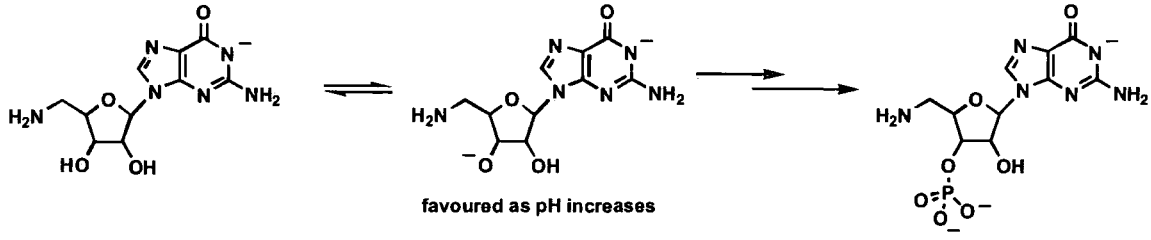


Fig: 2.57 We hypothesise the proportion of Amino-G deprotonated at the 2′, or 3′ hydroxyl to form the more nucleophilic oxyanion increases with pH. Which we hypothesise results in an increase in the yield of the phosphoester impurties.

The close structural similarity between these phosphate esters and the desired phosphoramidate would make separating these impurities from our product, whether by precipitation or gel filtration, difficult. Optimization of the reaction aided this process by minimising the amount of impurity present in the crude product.

2.55 Purification of GANP

At this point we determined that the crude GANP resulting from the aqueous reaction mixture was of sufficient quality to be used in a range of investigations. One of these studies was undertaken to assess the rate at which GANP's phosphoramidate group hydrolysed at various pH levels (see Chapter 3), carried out using ^{31}P NMR spectroscopy. We also began analyzed the products transcription reactions using the impure GANP, with the aim of providing an indication of the levels of incorporation achievable with a purified GANP sample (see Chapter 5).

Although aiming to avoid chromatographic methods we felt that if purification could be achieved in one step *via* size exclusion chromatography a minimum amount of product would be lost. We used Sephadex G10 gel filtration media, as in our previous attempt to synthesise and purify GANP by other synthetic methods. We expected the filtration to be more successful in this case as the salt and phosphate by-products of the reaction are of much lower molecular weight than GANP. To maintain a high pH a potassium carbonate eluent was used. A flow-UV detector was again used to detect elution of products.

The chromatogram showed two moderately resolved UV active peaks; the fractions relating to these signals were pooled and freeze-dried to give solid containing a high proportion of salt. However, although only one peak showed a phosphoramidate signal in the ^{31}P NMR spectrum, both peaks contained phosphate. Unfortunately signals in the ^{1}H NMR spectrum were very weak and therefore we were unable to ascertain whether the product had been separated from all the un-reacted starting material.

Through lowering the flow-rate and amount of crude material loaded onto the gel filtration column we hoped to improve peak resolution, thereby separating phosphate from the GANP. However, these alterations were to no avail as we consistently observed other phosphorus containing species along with the GANP peak trace. With gel filtration proving inadequate we looked for other methods of removing inorganic phosphate from the crude product.

2.56 Precipitative methods of barium ion removal

Seeing the removal of residual phosphate as the main challenge of the purification we sought an alternative work-up procedure, namely the precipitation of residual phosphate from the crude reaction mixture as a barium salt. By controlling the concentration of barium ions added to the crude solution we hoped to prevent the precipitation of our product as an insoluble salt, thereby avoiding the issues we faced previously when using barium ions (Section 2.34). We decided to use barium hydroxide as our metal ion source as in this way we could maintain the high pH conditions necessary to stabilise GANP.

This method of purification relies on the barium salt of inorganic phosphate being less soluble in aqueous solution than the barium salt of GANP.

By collecting the ^{31}P NMR spectrum of our crude reaction mixture we estimated the percentage of $POCl_3$ converted to GANP. This, combined with our knowledge of how much $POCl_3$ we had added to the reaction mixture, allowed us to calculate the amount of inorganic phosphate present in the crude mixture and add barium hydroxide accordingly. The white precipitate that formed was removed by centrifugation; a second ^{31}P NMR spectrum of the remaining solution showed the phosphate signal had decreased in intensity.

To remove trace amounts of barium from the sample we passed the solution through a Dowex ion-exchange resin (50 × 8W, Na^+ form) using a 1 mM NaOH eluent. All fractions exhibiting a UV trace were pooled and freeze-dried. This gave us our phosphoramidate product as a sodium salt in the presence of NaCl, NaOH, amine starting material, a reduced level of phosphate, and traces of the suspected 2′,3′ - phosphate esters.

We attempted to selectively precipitate the phosphoramidate through addition of 2.2 volumes of absolute ethanol. This is a standard method for the purification of nucleotides as precipitated sodium salts. The addition of the organic solvent coupled with cooling the solution to –20 °C caused the precipitation of a white solid, which was then collected by centrifugation. The ^{31}P NMR spectrum of the collected solid indicated the presence of both phosphoramidate and inorganic phosphate. We were unable to interpret the ^{1}H NMR spectrum, as the signals collected were too weak, being mostly obscured by the large amount of water present in the sample. Repeated precipitations did not improve matters, as our novel nucleotide seemed to display solubility in ethanol/water solutions too close to that of inorganic phosphate. The outcome of the repeated precipitations was a decrease in the amount of solid isolated upon each precipitation, but no change in the ratio of GANP to impurities.

2.57 Strong anion exchange chromatography

With the failure of both gel filtration and precipitation methods we opted to explore more time consuming chromatographic procedures. A study into the rate of hydrolysis of the phosphoramidate group of GANP (Chapter 3) showed that the pH at which Sephadex A25 is functional (pH 7 – 9) was within the range at which the our phosphoramidate product would hydrolyse while still on the column. Therefore we invested in a strong anion exchange media, Capto Q, which separates mixtures in the same manner as Sephadex A25 (competition for cationic binding sites) but is still functional at higher pHs. This means our product could be purified while still in its stable dianionic form. As Capto Q is an agarose based media it does not swell and contract under changing salt concentrations in the same manner as a Sephadex support. In terms of the practicality, the stable nature of the media means that Capto Q columns do not have to be re–poured after every use unlike when using Sephadex type media. This means that a single column of set length and width can be used to develop a standard method of eluent flow–rate and gradient. The use of a standard method allows us to be confident in comparing chromatograms resulting from the purification of a range of crude reaction phosphorylation reaction products.

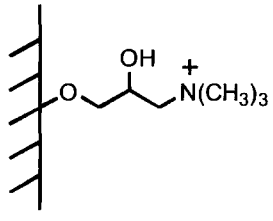


Fig:2.58 Chemical structure of Capto-Q chromatography support. Ammonium ion gives Capto-Q its functional properties through electrostatic binding to anionic molecules.

As with weak anion exchange media interruption of binding interactions and elution of anionic molecules is achieved through increasing salt concentration in the elution buffer. Compounds that are more anionic are bound to the support *via* stronger interactions than less anionic species. Dissociation of these strong interactions occurs at higher salt concentration than the weak interactions, resulting in variable elution times for compounds based on their charge alone. We hoped that the strong anion exchange

chromatography would allow us to separate our desired phosphoramidate from the amine starting material, inorganic phosphate and hopefully traces of the phosphate ester by-products. Unfortunately removal of residual salt in the crude reaction mixture is not achievable by anion exchange chromatography as the eluent contains a high salt concentration as a necessity. We intended to remove the salt eluted with the UV active GANP fractions by precipitation methods, expecting more success than previously experienced once the inorganic phosphate impurity had been removed.

Our standard method for chromatography with Capto Q utilized a $NaOH_{(aq)}$ eluent operating at 0.1 – 1 M concentration, which acted to both increase competition for anion binding sites and maintain a high pH. By using a sodium salt we aimed to simplify subsequent desalting steps. The resulting chromatogram showed good resolution of UV active substances as shown below. Fractions corresponding to each peak were lyophilized then redissolved in the minimum amount of water. We then used ^{31}P, ^{1}H NMR spectroscopy to determine how successful the separation had been.

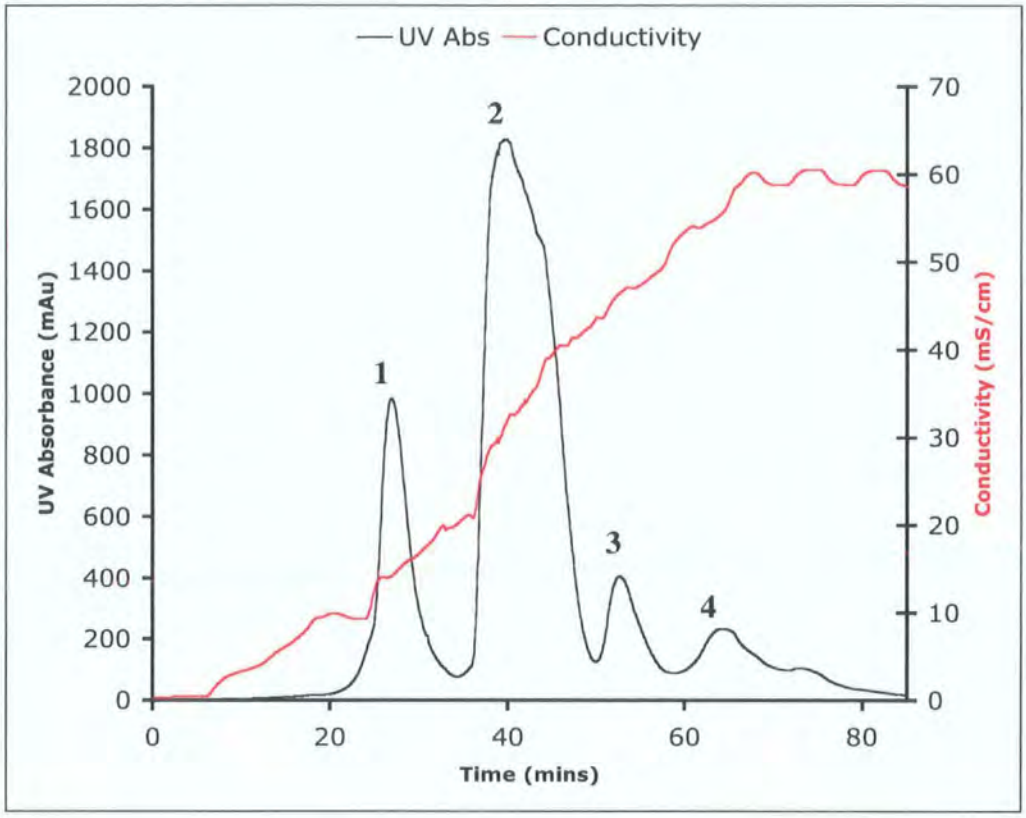


Fig:2.59 Plot of UV absorbance and Conductivity vs Time (mins). The chart shows well-resolved separation of reaction products according to their anionic charge.

Peak	^{31}P NMR signal	Assingment
1	6.6(s)	Phosphate
2	9.9(t)	GANP
3	7.6(s), 5.2(d), 2.79(d)	Phosphate and Phosphoesters
4	None	None

Table:2.6 Showing order of elution of phosphorus containing compounds.

The First compounds to be eluted are the neutral (or possessing a low negative charge) Amino–G (peak 1), which is the species responsible for the UV signal, and a trace of inorganic phosphate. The more anionic GANP and phosphate monoester groups elute close together later in the run owing to their stronger interactions with the Capto-Q media. The similar retardation of the two species can be attributed to their structural similarities.

2.58 Removal of excess salt from GANP sample

As the table above shows complete separation of GANP from other phosphorus containing species was achieved. We now faced the problem of removing excess NaOH from the sample. We were, again unable to remove this impurity *via* ethanol precipitation, collecting only GANP with a large quantity of salt. Repeated precipitations only resulted in a much-reduced yield of product for minimal reduction in salt levels. We decided an alternative method of salt reduction was required. As we no longer had the problem of separating our product from phosphate, we felt gel filtration was worth another try.

In the past we had failed when using this technique to purify crude reaction mixtures from inorganic phosphate. However, we felt more success could be had using G10 media as a method for removing excess salt from an otherwise pure sample. Initially we used a 1 mM NaOH eluent, aiming to reduce the salt present in the sample while maintaining high pH conditions. Upon examining the initial chromatograms it was apparent the GANP was only on the column for ~ 20 mins. We judged that owing to the relatively short time frame involved we could use a water eluent, thereby greatly reducing the salt concentration of the elutions, while a very small and hopefully undetectable level of hydrolysis would occur. The rate of hydrolysis would be minimised by both the phosphoramidate acting as a buffer in the otherwise neutral solution and the presence of residual salt from the NaOH fractions.

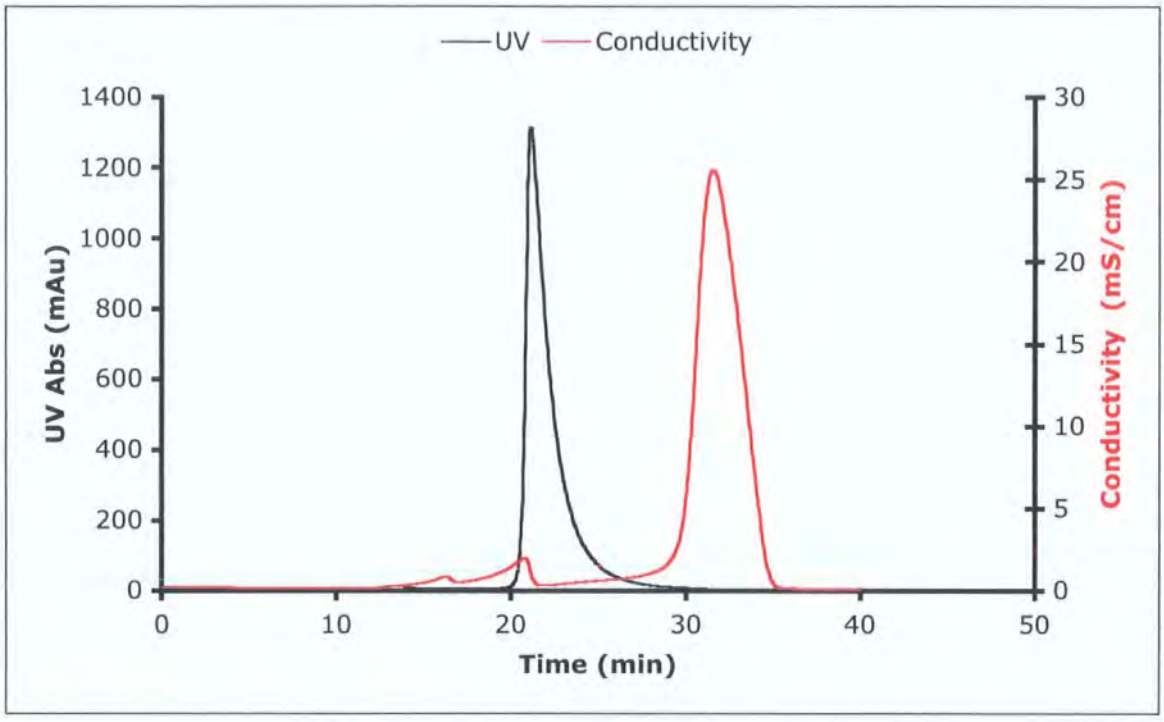


Fig:2.60 Plot of UV abs and Conductivity vs Time. Data for chart from gel filtration of GANP containing fractions eluted from the Capto Q ion exchange column. The plot clearly shows separation of the UV active GANP from the highly conductive inorganic salts.

As can be observed in the chromatogram above, separation of the UV active phosphoramidate and the highly conductive salts was very successful. Fractions relating to the UV signal were collected and freeze dried to give an off–white solid, which was subsequently dissolved in the minimum volume of water. The material still contained a small amount of NaOH from the gel filtration eluent (determined by pH). We were able to remove the majority of this salt through selective precipitation with ethanol.

The addition of two volumes of ethanol to the crude GANP solution resulted in the formation of a white precipitate, cooling the solution to –20 °C maximised the yield of solid. The solid was separated from the remaining ethanol: water solution by centrifugation. The collected solid was then washed with cold 70 % ethanol before being dried over phosphorus pentoxide under reduced pressure to give a pale yellow solid, which was characterised by ^{1}H, ^{31}P NMR spectroscopy, and high resolution mass spectroscopy.

2.59 Characterisation of GANP

The ^1H NMR spectrum of the isolated solid revealed the presence of residual water, obscuring the signals. To remedy this we displaced the water by repeatedly dissolving our product in D_2O then lyophilizing the resulting solution. This allowed us to acquire far clearer ^1H NMR spectra of GANP.

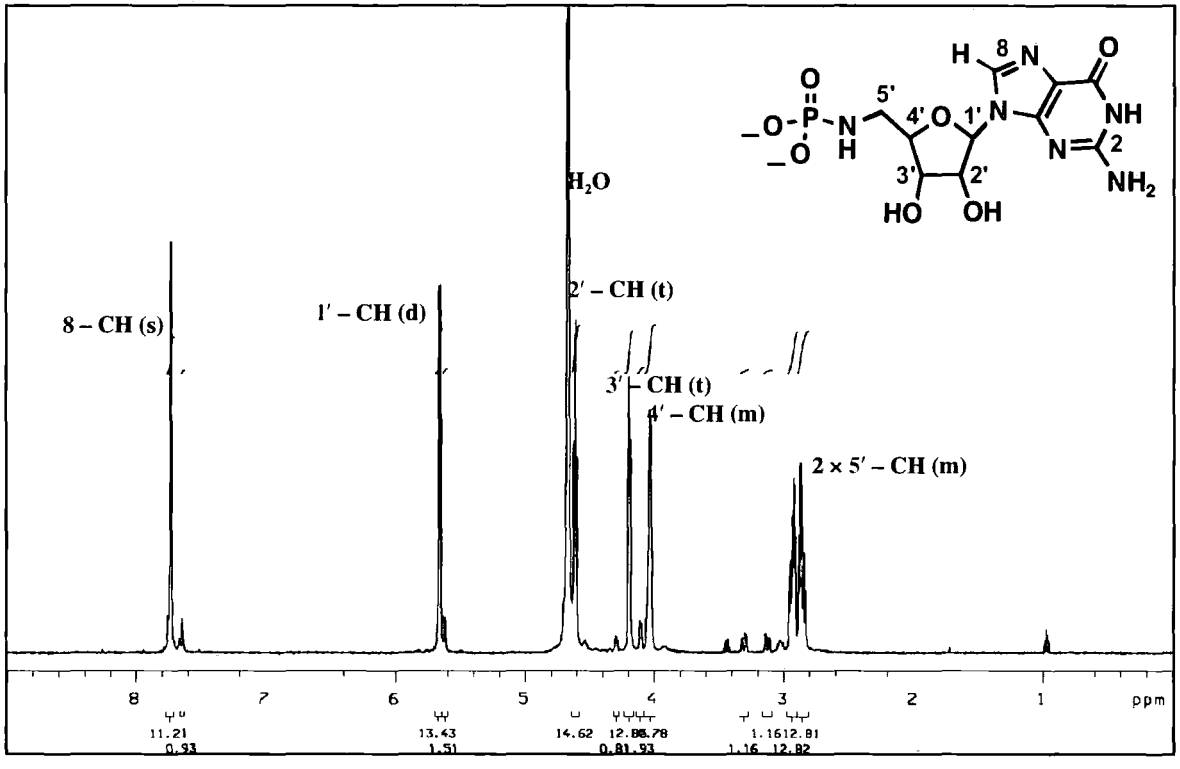


Fig:2.61 ^1H NMR spectra of GANP in D_2O (20 °C, 500 Mhz).

We can assign all the major peaks in the spectrum to protons in various environments present in GANP. However it is apparent that a number of unassigned low intensity signals are also present. Through comparison to the ^1H NMR spectra of Amino–G we can provide an argument for the majority of these signals being due to Amino–G, present following hydrolysis of GANP. The remaining impurity signals can be assigned to residual ethanol present in the sample. The Amino–G peaks do not overlap with those of GANP, and therefore do not inhibit the assignment of peaks in the ^1H or ^{13}C NMR spectra.

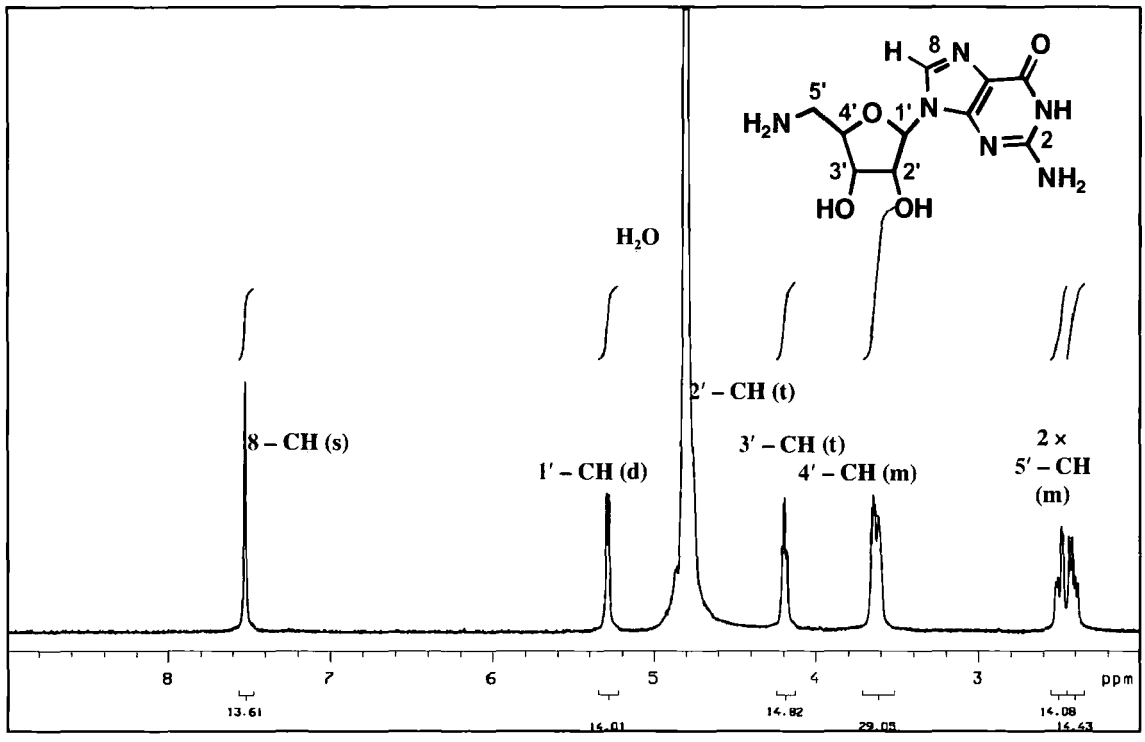


Fig:2.62 ^{1}H NMR spectra of Amino–G in 4 % NaOD:D_2O (v/v) (20 °C, 400 Mhz).

The chemical shifts and splitting patterns for Amino–G are very close to those of GANP. The similarity of the two species NMR spectra gave us a clue to the source of the low intensity signals present in the GANP ^{1}H NMR spectrum.

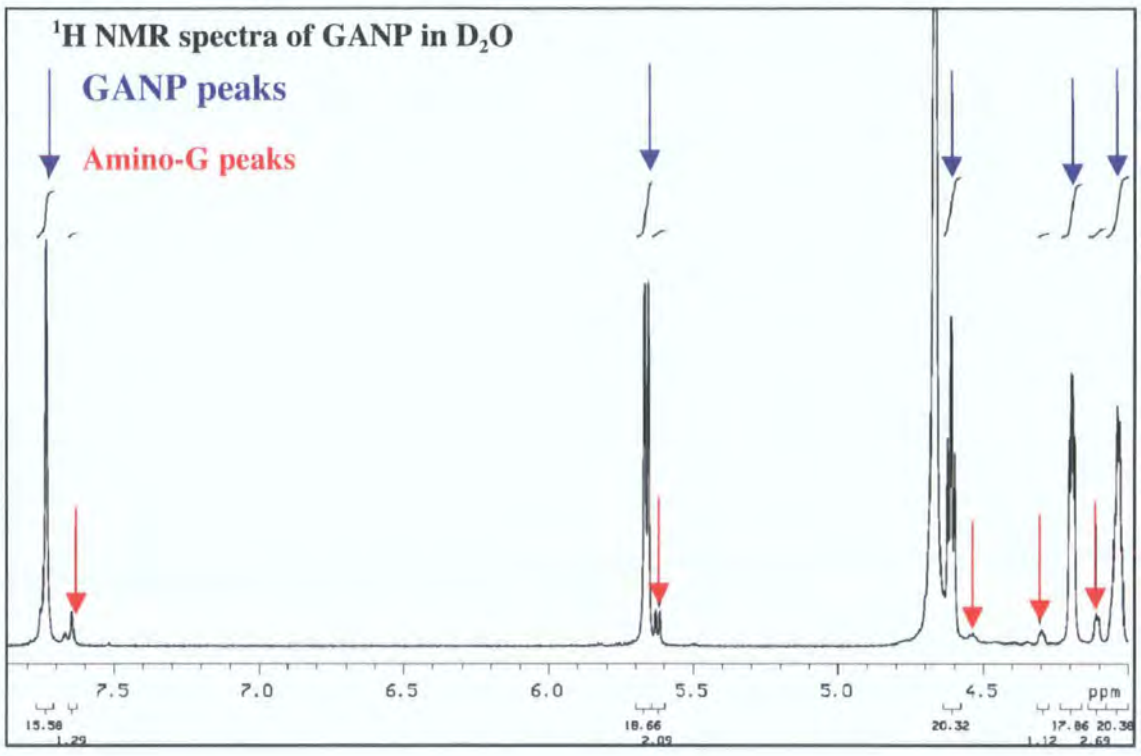


Fig:2.63 ^{1}H NMR spectra of GANP in D_2O (20 °C, 500 Mhz). This section of the spectra has been used to highlight the correlation between the higher intensity GANP signals (blue arrows) and the lower intensity Amino–G peaks (red arrows).

Impurity peaks in GANP ^{1}H NMR spectra (other than those assigned to ethanol) show correlation to GANP signals of similar chemical shift and coupling. The similarity between the shifts and coupling patterns of the two sets of signals indicates they share a common structure. We believed that these low intensity peaks indicated the presence of Amino–G in the sample, either from the unlikely case that amine starting–material had been carried through the purification process, or more believably as a product of GANP hydrolysis. Through analysis of the characteristic signal for the 5′-CH_2 of Amino–G and GANP in ^{1}H NMR spectra we were able to corroborate this assignment.

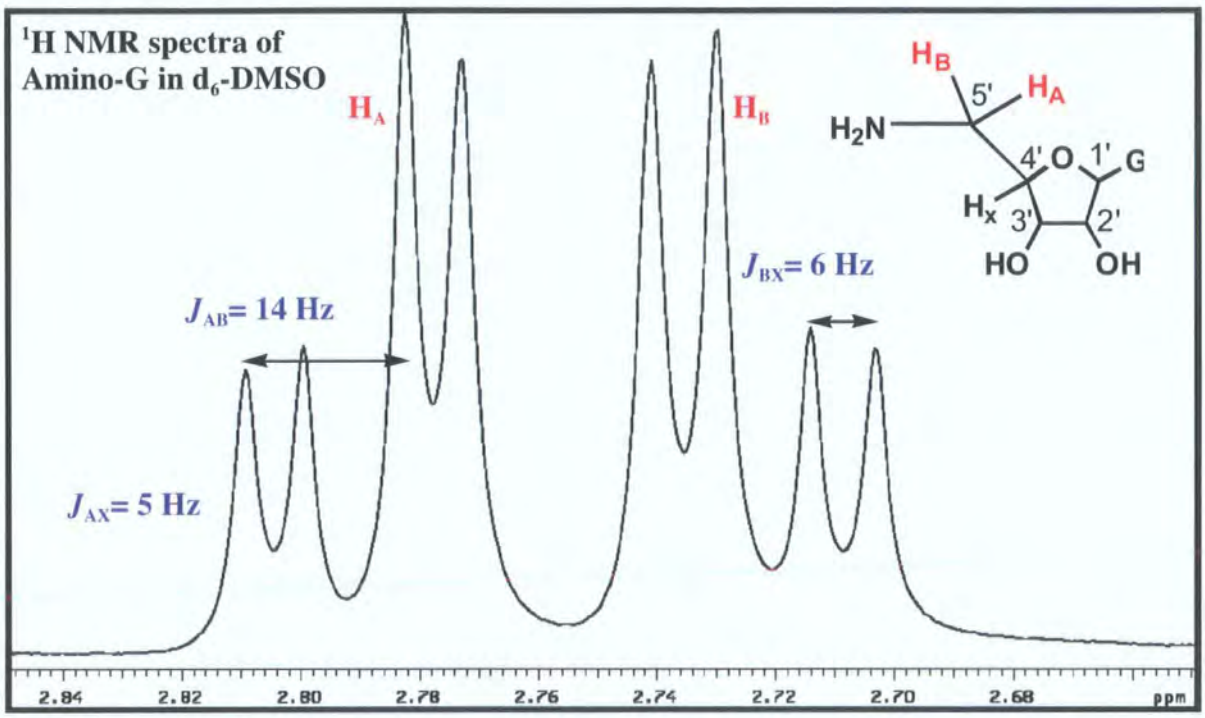


Fig:2.64 ^{1}H NMR spectra of Amino–G in d_6-DMSO (20 °C, 400 Mhz) showing detail of the 5′-CH_2 signals.

Analysis of the 5′-H signals in the ^{1}H NMR spectra of Amino–G reveals an ABX coupling environment. Each proton attached to the 5′- carbon is in a unique chemical environment, resulting in two separate, though very close, chemical shifts. Hence, each 5′-$H_{A/B}$ proton signal displays ^{3}J–vicinal coupling not only to the 4′-H_X but also couples (^{2}J) with the remaining geminal 5′- $H_{A/B}$ proton. These coupling characteristics result in each proton displaying a doublet-of-doublets splitting pattern. The near identical environments of the two 5′- $H_{A/B}$ has the added effect of producing a degree of 'roofing' between the two signals.

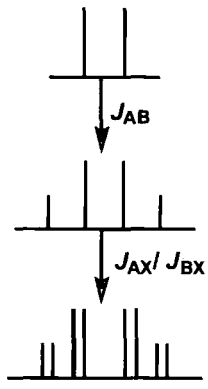


Fig:2.65 Diagram of the proposed splitting pattern of the 5′–H_AH_B protons of Amino–G.

The largest coupling is that between the H_A and H_B protons and is measured from the outer most peaks of each doublet signal (J_{AB}/J_{BA} = 14 Hz). The coupling between the methylene protons and the .4′- H_X is less strong, possibly owing to the distance from, or angle between, the 5′- $H_{A/B}$ protons. The magnitude of the $J_{AX/BX}$ coupling varies for each 5′- proton owing to their differing levels of interaction (J_{AX} = 5 Hz, J_{BX} = 6 Hz). If the 5′- CH_2 protons were not diastereotopic we would expect to observe the 4′- H_X signal as a quartet in the ^{1}H NMR spectra. However, as the H_A H_B protons do not couple to the H_X proton in an equivalent manner, the 4′–H_X signal has a more complex multiplet splitting pattern.

In the ^{1}H NMR spectra of our final GANP product we have assigned signals in the range 2.7 – 3.4 ppm as the same 5′- $C(H_AH_B)NH_2$ ABX system as present in the Amino–G spectra, albeit in low intensity. This ABX system is very characteristic and we view its presence in the spectra as key in assigning the impurity signals to Amino–G, the source of which will be discussed below. A significant feature in the GANP spectra was a multiplet at ~ 2.9 ppm that appeared to share features with the ABX system discussed above, yet appeared to be more complex.

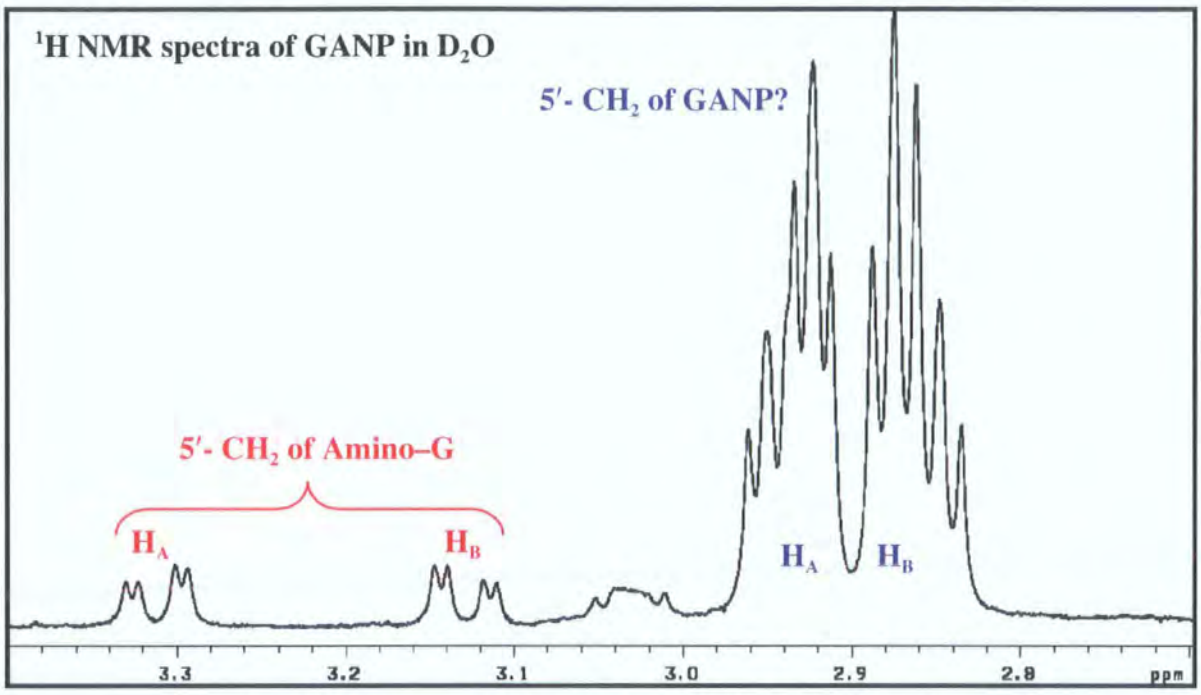


Fig:2.66 ^{1}H NMR spectra of GANP in D_2O (20 °C, 500 Mhz) showing the presence of both the 5′-proton signals of Amino–G in their ABX system, and the similarly coupled 5′-protons of GANP.

The most likely source of the multiplet signals at ~2.9 ppm is the 5′-CH_2 group of GANP. To confirm this assignment though we must first explain the splitting pattern when compared to that observed for the 5′-CH_2 signals for Amino–G. The most significant difference in the environment at the 5′ positions of GANP and Amino–G is the introduction of a phosphate-containing group. As we have previously mentioned ^{31}P is the only significant natural isotope of phosphorus, and is NMR active. We have already observed coupling between the phosphorus and methylene protons in the ^{31}P NMR spectra of GANP, resulting in a triplet splitting pattern. The interaction between the methylene protons and the phosphorus atom results in a splitting pattern that differs from the ABX system observed in Amino–G, though it does have a similar basis.

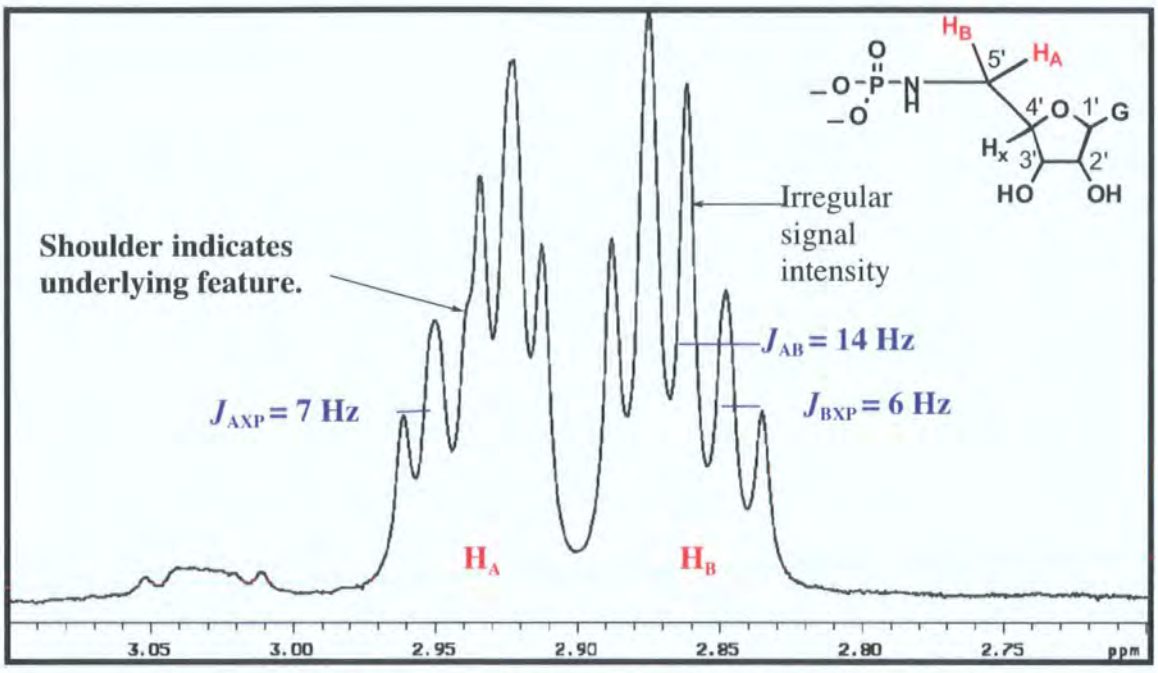


Fig:2.67 ^{1}H NMR spectra of GANP in D_2O (20 °C, 500 Mhz), showing the 5′–H_AH_B signal.

In a fashion similar to that observed for the 5′–methylene protons of Amino–G, the 5′-methylene of GANP arise as a complex signal in the H^1 NMR spectrum. As discussed earlier this coupling pattern is owing to the two protons being in differing chemical environments, combining to form an AB splitting system. Each proton couples to the other methylene proton splitting each proton signal to form a doublet that displays 'roofing' in the same manner as the Amino–G 5′–H_AH_B signal, resulting in an identical coupling constant (J_{AB} = 14 Hz). Subsequent coupling to both the 4′-H_X and phosphorus atoms splits each doublet peak into what appears to be close to a triplet signal (J_{AXP} = 7 Hz, J_{BXP} 6 Hz) in a type of ABX system. The coupling constants are each 1 Hz larger than that of the respective J_{AX} and J_{BX} values of the 5′–H_AH_B signals of Amino–G. By collecting a ^{1}H NMR with ^{31}P decoupled each proton signal can be seen to revert to an Amino–G like ABX doublet of doublets, with J_{AX} = 6 and J_{BX}= 5, confirming the source of the additional coupling as the phosphorus of the phosphoramidate group (Fig:2.69).

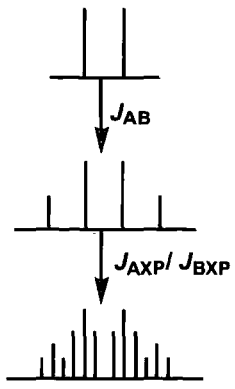


Fig:2.68 Diagram of the proposed splitting pattern of the 5′–H_AH_B protons of GANP.

Poor resolution of the 5′- proton peaks in the GANP spectra makes assignment of a triplet splitting–pattern tentative at best. However, the 'shoulder' present on the peak at ~2.96 ppm indicates an underlying peak belonging to the outer H_A triplet set. This in conjunction with the intensity ratios of the signals convinced us a doublet-of-triplets splitting pattern was the correct assignment. The lack of a 'shoulder' on the complementary H_B peak is a result of a closer overlap between the inner and outer sections of the split signal, as indicated by a higher than expected intensity observed for this peak. The splitting patterns for H_A and H_B do not mirror each other owing to the differing environments in which the two protons exist.

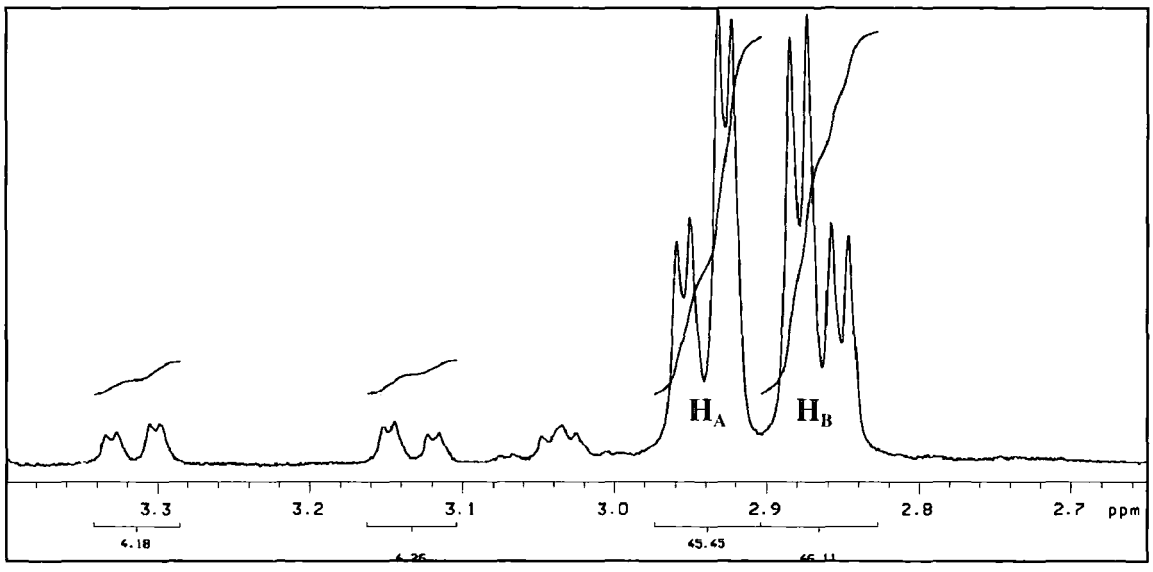


Fig: 2.69 ^{31}P decoupled ^{1}H NMR spectra of GANP in D_2O (20 °C, 500 MHz). By removing the ^{31}P coupling we can observe how the complex 5'-H_AH_B signals of GANP revert to an Amino–G like ABX system. This confirms that the additional splitting evident in Fig: 2.40 arose as a result of coupling to ^{31}P.

Further evidence to support the successful synthesis of GANP was gained from both the ^{13}C and ^{31}P NMR spectra.

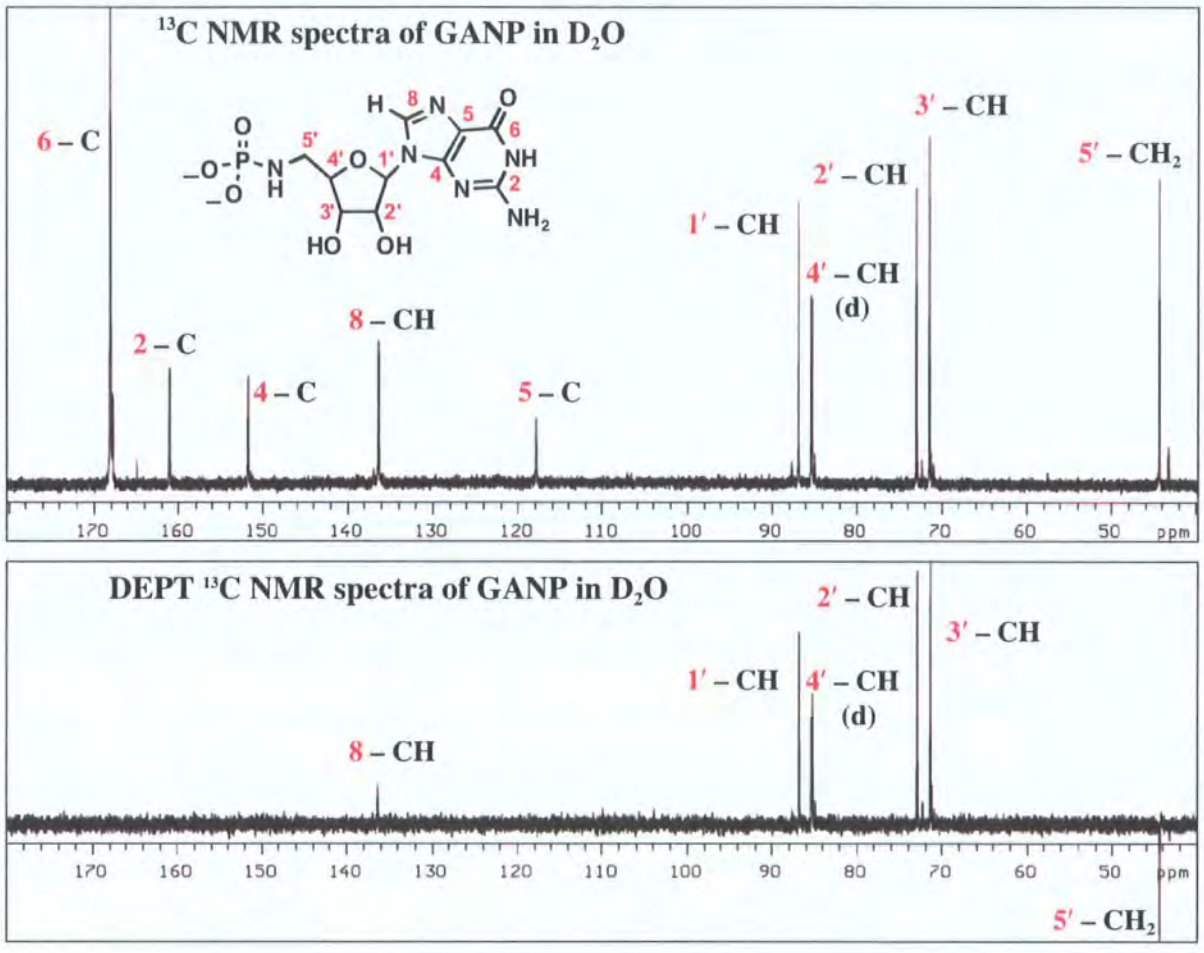
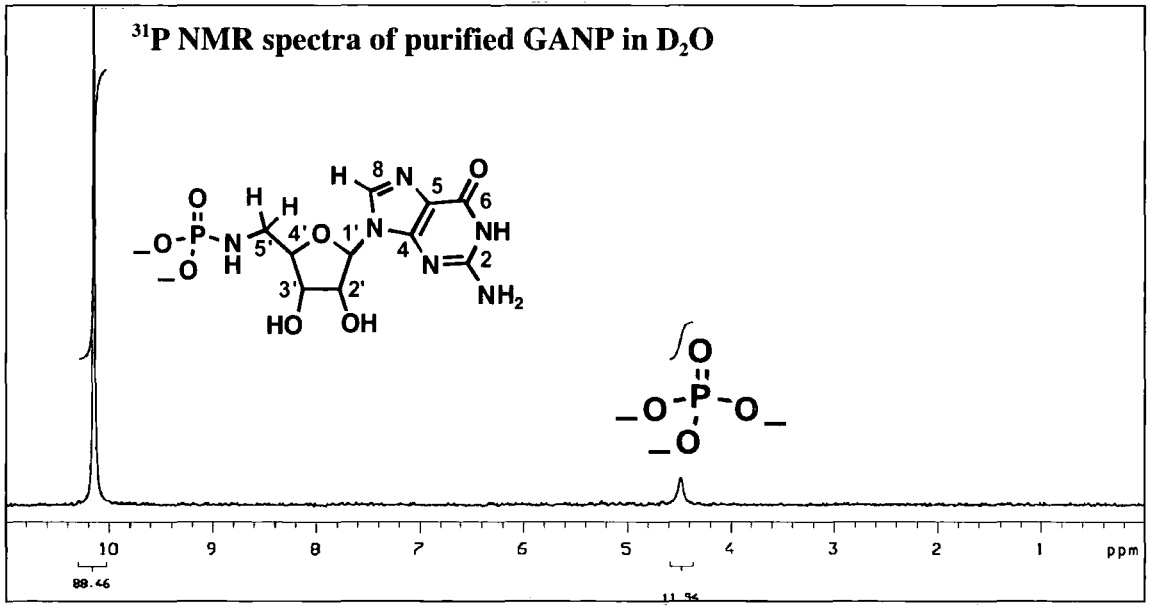


Fig: 2.70 ^{13}C and DEPT spectra of GANP in D_2O. Peaks assigned according to chemical shift. The presence of coupling to ^{31}P adds further evidence to a successful synthesis.

All the major peaks in both the ^{13}C and DEPT spectra were assigned to GANP. In corroboration with the ^{1}H NMR spectrum we could observe the presence of Amino–G through low intensity peaks at chemical shifts close to those of some GANP signals. The most visible of these impurity peaks can be found close to the 5′–C peak of GANP (44.40 ppm) and was assigned to the 5′–carbon of Amino–G. As phosphorus is NMR active, we looked to the 5′–C signal for signs of coupling. Unfortunately no splitting of the signal was visible, this is entirely reasonable as the range of $^{2}J(^{13}C,^{31}P)$ coupling runs from –20–+50 Hz, making a coupling constant close to zero possible. In the absence of ^{2}J coupling we analyzed the other carbon signals for splitting patterns. The only peak that showed coupling was the 4′–C signal, which was split to a doublet through $^{3}J(^{13}C^{31}P)$

coupling of 10 Hz. The coupling constant is again entirely reasonable, and though we would have preferred to observe both ^{2}J and ^{3}J coupling the presence of any $^{13}C^{31}P$ coupling provides more evidence for the success of our synthetic route and confirms the site of phosphorylation.

Throughout our attempts to synthesise GANP we have relied upon ^{31}P NMR as our primary means of analyzing our crude reaction products, and to monitor our attempts at their purification.



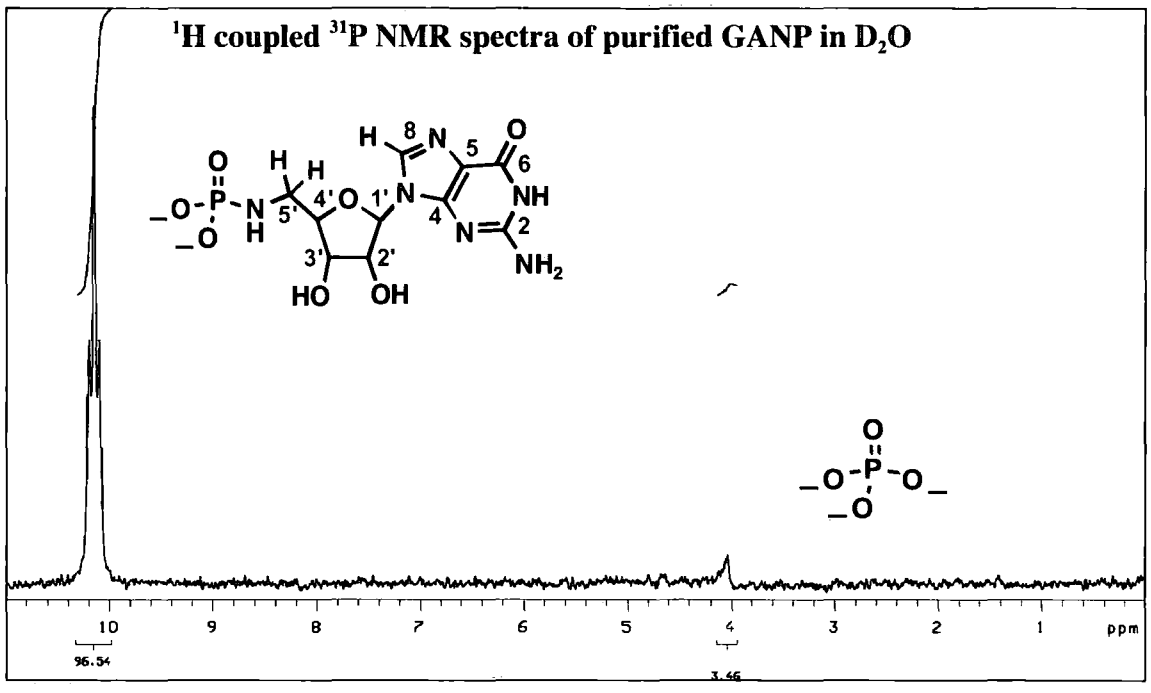


Fig: 2.71 The two ^{31}P NMR spectra both show a the phosphoramidate peak of GANP in the presence of a small quantity (~10 %) of inorganic phosphate. Coupling to the two 5′-methylene protons of GANP produces a triplet signal. Phosphorylation at any other position on Amino–G would result in either singlet or doublet signals.

As the spectra above show our final GANP sample contained two phosphorus species. Our phosphoramidate product, appearing at ~ 10.2 ppm, was the most abundant of these contributing to 90 % of the phosphorus containing species in our GANP sample. Assignment of this signal was made its appearance as a triplet in the proton coupled ^{31}P NMR spectra, $^3J(^{31}P,^1H) = 6$ Hz.

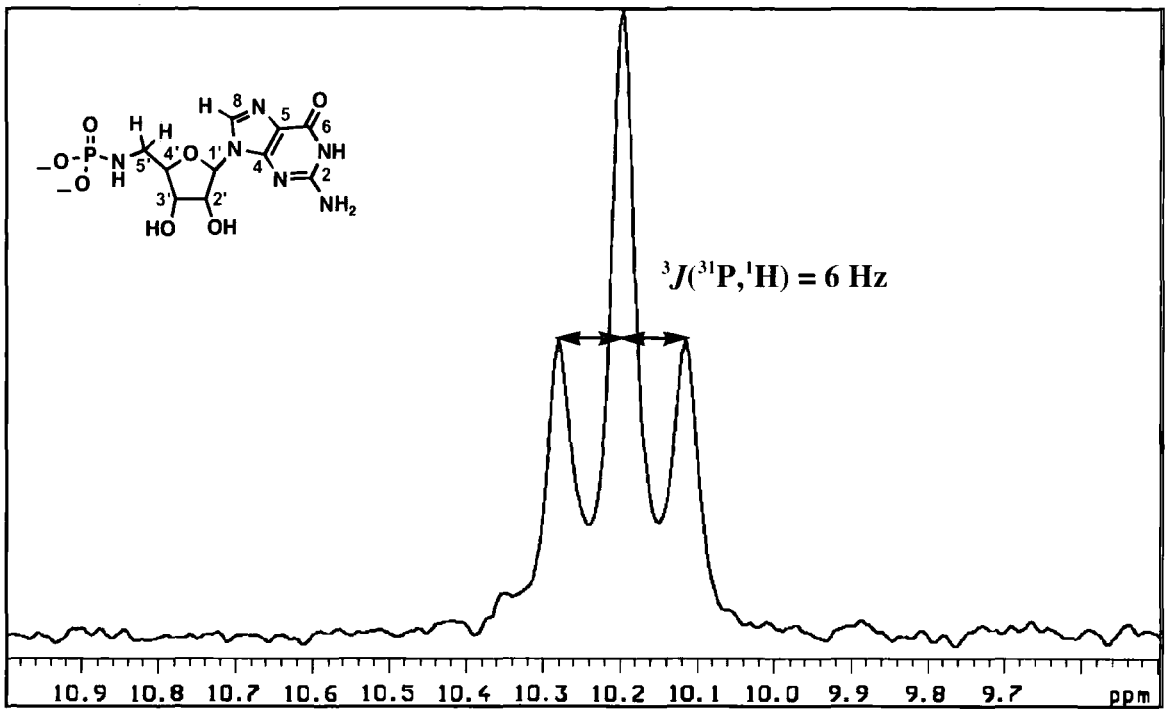


Fig: 2.72 ^1H coupled ^{31}P NMR of GANP in D_2O.

The second signal, assigned to the inorganic orthophosphate ion (10 %), had a chemical shift of ~ 4.5 ppm. It is possible that both inorganic phosphate and amine starting material could be carried through the work-up procedure and remain present in the final product. However, owing to the nature of the chromatographic techniques employed it seems extremely unlikely that this would occur. A more believable hypothesis is that during the multiple precipitation steps of GANP purification a degree of hydrolysis of the phosphoramidate occurred.

In addition to NMR data we also collected a high-resolution mass spectrum of the GANP sample. The collection of accurate elemental analysis data was made impossible owing to the hygroscopic nature of the solid and its ambiguous ionisation state.

2.60 Assessment of the synthetic route

Although we maintained high pH conditions where possible throughout the work-up steps, we were unable to use high salt conditions during our final desalt chromatography column. Attempts to use even a low salt buffer meant our eluted product, when freeze-

dried and redissolved in a low volume of water, contained a high salt concentration (assessed by pH), making the remaining precipitation steps unworkable. As we were aware some hydrolysis was possible during this stage the chromatography was undertaken in the minimum amount of time possible and the products were frozen as soon as they were eluted. Even so it is apparent some hydrolysis did occur which serves to attest to the lability of phosphoramidate groups in aqueous conditions, and adds weight to the presence of a phosphoramidate.

The degree of hydrolysis, and subsequent presence of Amino–G and phosphate impurities in the sample, although not ideal should not inhibit our ability to use the collected GANP in transcription reactions. The presence of a small amount of phosphate is of no consequence considering we expect at least 50% of the GANP entered into the transcription reaction to hydrolyse over the 2.5 h reaction time. As phosphate species (specifically pyrophosphate) are by-products of the transcription reaction we expect presence of these phosphate groups to have little to no effect on reaction products.

As we know the proportion of Amino–G present in our sample we can determine stock solutions of GANP for use in transcription by UV spectroscopy, factoring in absorbance owing to Amino–G. As we will demonstrate in a subsequent chapter the presence of Amino–G does not have a negative effect on the transcription reaction. Also it will be present in the reaction as a by-product of GANP hydrolysis meaning the small amount present at the beginning of the reaction will be negligible when compared to the larger quantity formed as the reaction proceeds.

3.0 GANP hydrolysis study

3.1 Aims of the study

Prior to studying the ability of 5′-amino-5′-deoxyguanosine 5′-N phosphate (GANP) to initiate RNA synthesis we felt it necessary to determine the phosphoramidate group's stability under transcription reaction conditions. As mentioned in previous chapters phosphoramidates are subject to hydrolytic decomposition, the rate of which is pH dependent (Section 2.3). Assuming our hypothesis is correct, the presence of an anionic phosphorus moiety at the 5′- position is key for the initiating ability of GANP. If the phosphoramidate is unstable under transcription conditions the effectiveness of GANP could be greatly compromised. For the sake of completeness we intended to not only observe the hydrolysis under transcription conditions but also over a range of pH levels, we will then use the data acquired to create a pH rate profile for the reaction. Past studies of phosphoramidates have shown the rate of hydrolysis to be faster at low pH ($t_{1/2}$ in the order of minutes) and slower at high pH ($t_{1/2}$ in the order of days)[48-50]. Therefore we required a method of analyzing the rate of GANP hydrolysis over both short and long time periods, also as transcription reactions are undertaken at 37 °C we needed to be able to incubate the samples and carry out measurements above room temperature.

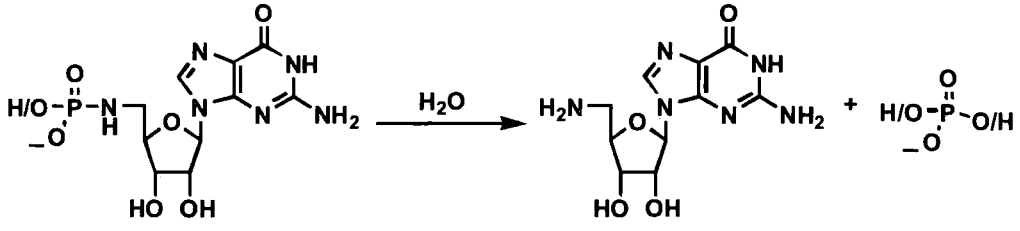


Fig:3.1 Hydrolysis of GANP

3.2 Method of analysis

The analytical technique we chose to employ was ^{31}P NMR spectroscopy. We had already confirmed during the synthesis of GANP that the phosphoramidate and phosphate species appear as distinct signals in the NMR spectrum. This allowed us to detect the relative amounts of both the starting material and product of the hydrolysis present in the

sample. By integration of the respective peaks in a spectrum we were able to determine the ratio of GANP: phosphate. By repeating the integration and ratio determinations for a number of spectra collected over a range of time points, we were able to determine the rate of change in the GANP signal, and hence the observed rate constant (k_{obs}) of hydrolysis at a specific pH. By incubating samples in differing pH levels under pseudo first order conditions and collecting NMR spectra over a series of time points for each we were able to collect the data necessary to construct a pH rate profile.

3.3 Purity of GANP used

We carried out this pH study before we had developed a satisfactory purification method for GANP synthesised *via* our aqueous phosphorylation reaction. Therefore the samples used to carry out the investigation were crude in that they still contained residual salt, amine starting material, and phosphate. As all the GANP samples used were sourced from the same crude reaction product the quantities of these impurities was constant in each reaction studied. The nature of the contaminants is such that, as they are neither acidic nor present in high quantities, they should not contribute to the rate of GANP hydrolysis. Each NMR sample contained the same quantity of GANP and had its pH adjusted with buffer before measurements began. The concentration of each buffer used was maintained as a constant (0.5 M) across the range of experiments carried out to avoid complications owing to the unlikely event of general acid catalysis occurring.

3.4 Acid catalysis considerations

It is known that acid catalysis can function *via* two mechanistic pathways, specific and general acid catalysis. In a specific acid catalysed reaction the rate of reaction increases linearly with $[H^+]$. The proton source in such reactions being the hydronium ion, H_3O^+, meaning the concentration of any buffer/weak acid (AH) used is unimportant. A general acid catalysed reaction shows a linear increase in the observed rate constant (k_{obs}) upon an increase in concentration of a weak acid, AH, at constant pH. Such reactions are catalysed by AH as well as H_3O^+, a quality that allows easy identification of which process is occurring in a specific reaction.

No published data has identified general acid catalysis occurring in phosphoramidate hydrolysis, therefore it was not our intention to probe or in any way comment on the nature of the acid catalysed hydrolysis of GANP. We were satisfied to maintain a constant buffer concentration across the experiments studied, thereby keeping any contribution to the rate of hydrolysis from a general acid catalyzed pathway constant across the samples and maintaining the comparability of the results. The table below gives the pH levels at which the hydrolysis was studied, and the buffer used in each case.

pH	Buffer/Base
3	Formate
3.5	Formate
7.2	MES
8	Bicarbonate
9	Borate
9.8	Carbonate
10.5	NaOH

Table:3.1 The table above shows the pH levels at which the phosphoramidate hydrolysis was observed, and the buffers or base we utilized to maintain the pH conditions.

3.5 Data acquisition

Our methodology for tracking the hydrolysis of GANP differed according to the pH of the reaction conditions. A number of methods were essential as the half-life of the hydrolysis reactions studied ranged from minutes to days, too extensive a range to use a single procedure.

3.5.1 Low pH conditions (3 – 3.5)

Under these conditions hydrolysis was at its fastest with a half-life on the order of minutes. As the acquisition of a clear ^{31}P NMR spectra using the amount of GANP we were employing in each experiment took approximately five minutes, it was not possible to use a single sample to follow the degree of hydrolysis over a set time period. Instead we set up a larger scale reaction at a defined pH (3, 3.5) and temperature (37 °C), and then

after predetermined periods of time, removed aliquots that were immediately quenched in base to halt the hydrolysis reaction. The ^{31}P NMR spectra of these quenched samples were then collected at our leisure, enabling us to assess the degree of GANP hydrolysis at each time point and thereby calculate the rate of hydrolysis.

3.5.2 Mid pH conditions (7.2 – 9)

The moderate $t_{1/2}$ of the reaction under these conditions made the mid pH range of experiments the most straight–forward degradations to track. The GANP sample was made up to the required pH in an NMR tube then incubated at elevated temperature in the spectrometer. We were able to program the spectrometer to make acquisitions at hourly time periods, enabling us to collect enough data to determine the k_{obs} for hydrolysis of each sample. An example of spectra collected using this methodology can be seen in the stack plot below, in which the reduction of GANP and increase of phosphate signals can be observed.

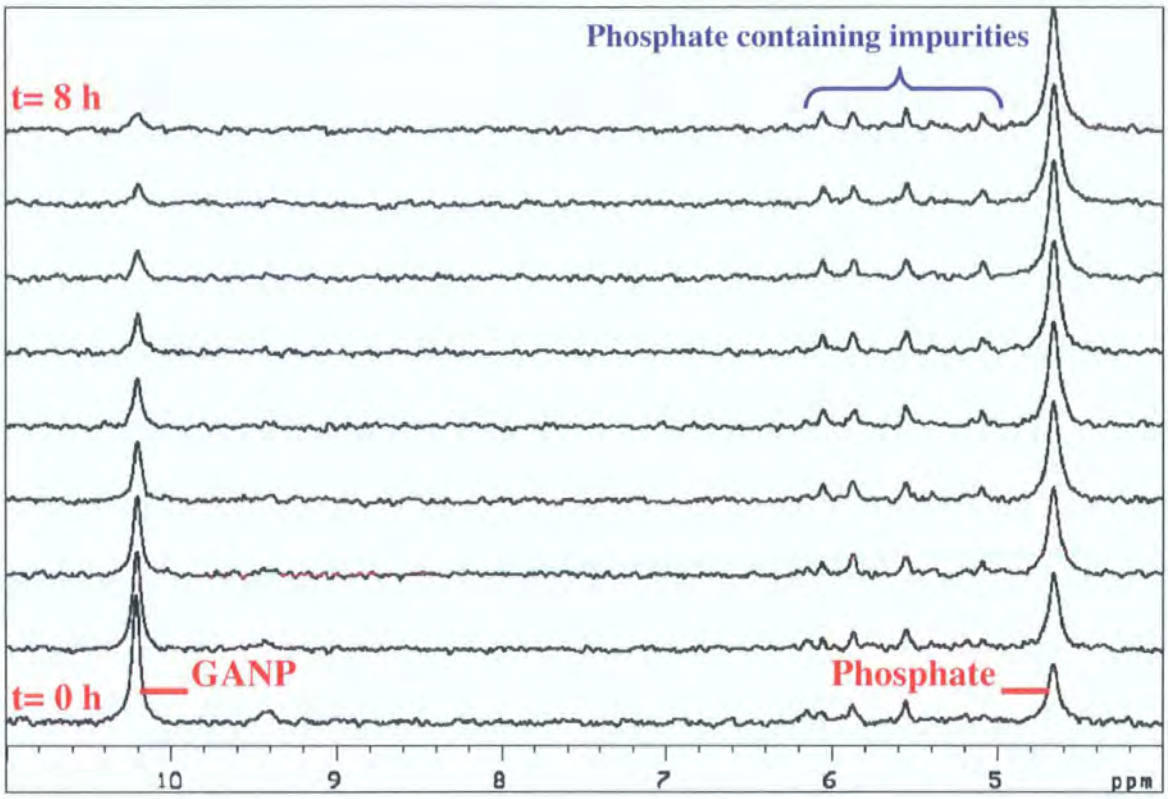


Fig:3.2 The plot above shows a representative set of ^{31}P NMR spectra collected over the course of 12 h (last four spectra not shown) from a sample held at 37 °C and pH 8. The degradation of GANP and formation of phosphate over time can be clearly seen in the increase and decrease of the respective peaks. Integration of the GANP peaks relative to inorganic phosphate in each of the spectra allowed us to quantify rate of hydrolysis as shown below.

3.5.3 High pH conditions (9.8–10.5)

As we needed to collect spectra of these samples infrequently owing to the slow rate of reaction, and it was unfeasible to run a NMR spectrometer at raised temperature for this length of time, we devised a protocol different from that used in lower pH experiments. The reaction mixture was made up in the NMR tube as in previous experiments with the pH level being maintained by a buffer. After the initial NMR spectra had been acquired, at room temperature as the rate of reaction was low enough to be unaffected by a slight temperature fluctuation, the sample was incubated in a water bath maintained at 37 °C. Subsequent spectra were collected at predetermined times before returning the sample to the water bath to continue the reaction.

3.6 Inherent limitations

Using these various methods we were able to derive k_{obs} values for the hydrolysis occurring in conditions of pH 3–3.5 and 7–10.5. Unfortunately we were unable to follow the hydrolysis in the pH range 3.5–7 as GANP was insoluble in water under these conditions. We hypothesised our inability to dissolve the GANP under these conditions was owing to the formation of a neutral zwitterion which when coupled with the notoriously insoluble guanine moiety resulted in the observed insolubility.

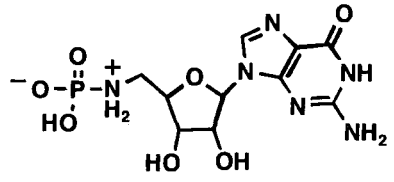


Fig:3.3 The insoluble GANP zwitterion present at neutral pH.

The presence of a mono/di-anion and mono-cation at higher and lower pH levels respectively was thought to facilitate the solvation of the phosphoramidate by counteracting the insoluble guanine base (Fig:3.4). The data we were able to acquire allowed us to construct the pH vs log k_{obs} plot shown below. Although obviously not as complete as we would have liked, it does allow us to determine the trend of the changing k_{obs} values. Most importantly we were able to determine the half–life of the phosphoramidate under transcription reaction conditions, in conjunction we also observed an increased rate of hydrolysis at lower pH levels that would facilitate the phosphate group's efficient removal.

3.7 Treatment of data

The following section will detail how the data collected from each observed hydrolysis was manipulated to calculate the respective rate constants. Rather than trawl through the manipulation of each data set, I will provide a generic example showing the steps we followed. The table below shows the collected peak integration data for the observed period of GANP hydrolysis at pH 8 (Fig:3.2).

Time (h)	GANP peak Integrals	Normalised Results
0	100	1
1	76	0.76
2	63	0.63
3	44	0.44
4	34	0.34
5	27	0.27
6	18	0.18
7	14	0.14
8	12	0.12
9	9	0.09
10	8	0.08
11	5	0.05
12	3	0.03

Table:3.2 Shows results for the hydrolysis of GANP at pH 8, as observed by ^{31}P NMR spectroscopy. (the table includes the 4 spectra not included in Fig:3.2). Results are normalised against the initial level of GANP to allow for simplified interpretation of the data.

Through manipulation of the data in Table:3.2 we were able to use a graphical method to determine the k_{obs} for each reaction that was followed. This method of data analysis was made possible by treating the hydrolysis reaction as a pseudo 1st order process (made possible by the large excess in which water was present), with the rate equation:

$$rate = k'[GANP]$$

Manipulation of this equation gives us the exponential form:

$$[GANP]_t = [GANP]_0 . e^{-k't}$$

For the data shown above (Table:3.2), using the exponential equation and plotting [GANP] against time gives us:

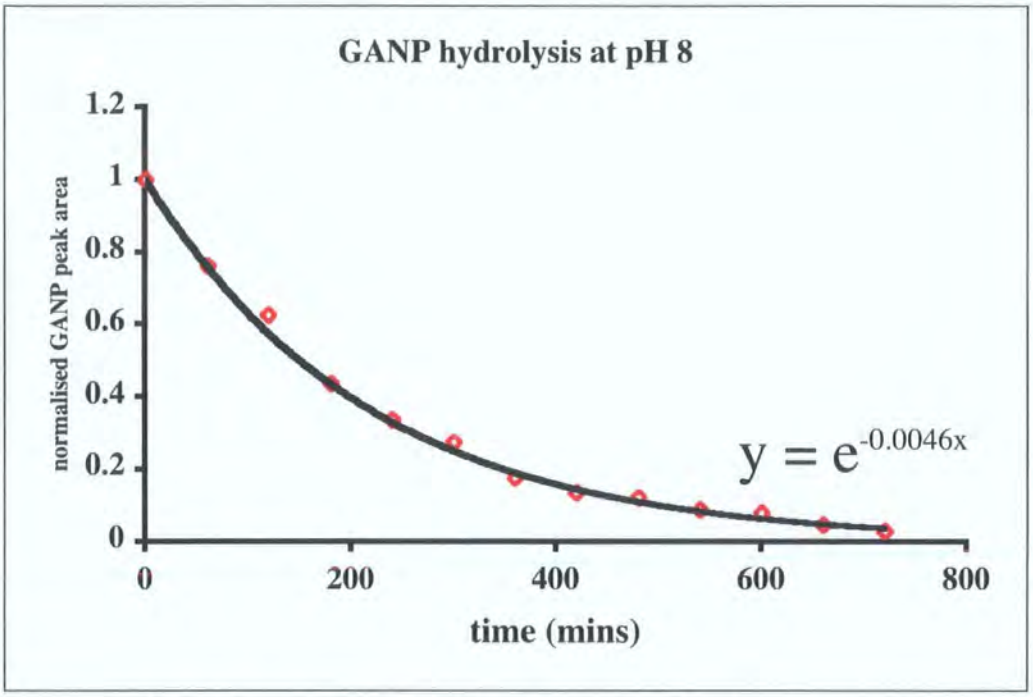


Fig:3.4 Plot of normalised pH 8 integrals *vs* time. Windows Excel software was used to plot an exponential line of best fit through the data points, the equation of which was used to estimate the rate constant of the hydrolysis.

Windows Excel software was used to plot an exponential curve through the experimental data points. The plot shows a good fit between the experimental data and the generated exponential trend–line, meaning we could use the equation of the trend–line to calculate the k_{obs} value for the hydrolysis. The pseudo 1st order rate constant ($k_{obs)}$ could then be used to derive the half-life of the hydrolysis at the pH studied as well as included in a pH rate profile.

3.8 pH vs k_{obs} plot

Through manipulation of the acquired NMR data we were able to derive the k_{obs} value at each pH level studied. This allowed us to both determine the half-life of the hydrolysis over the range of pH values, and create the pH rate profile shown below.

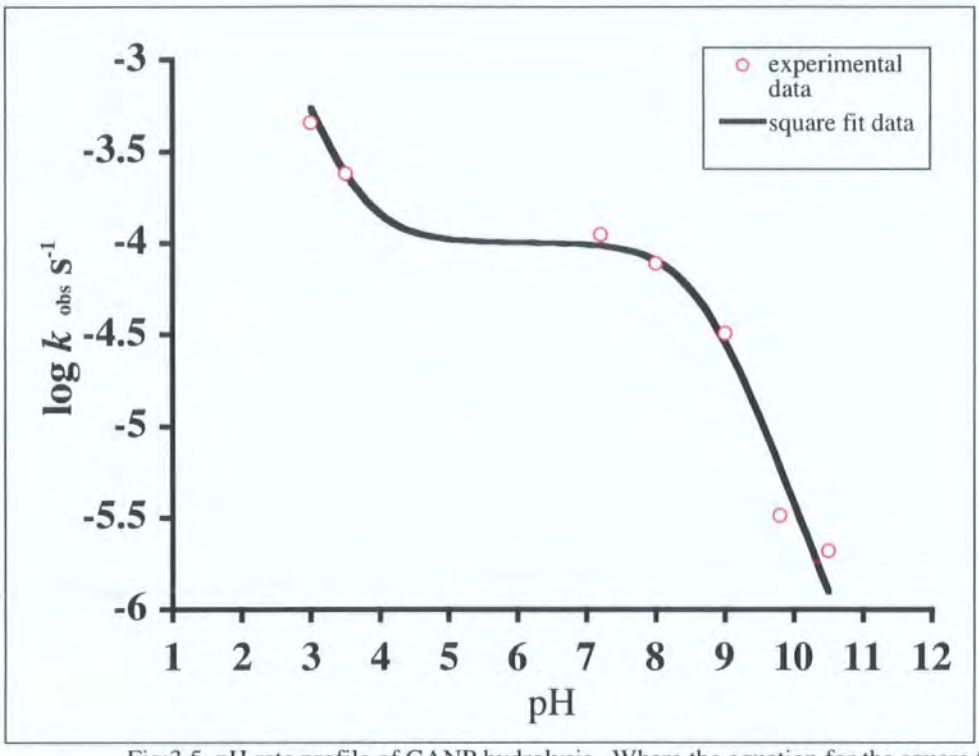


Fig:3.5 pH rate profile of GANP hydrolysis. Where the equation for the square fit data is: $k_{obs} = 26.7266 \times 10^{pH} + (0.00603573/ (1+10^{pH-8.599}))$

The plot shown above (Fig:3.5) features experimental k_{obs} values, determined by methods shown in Section 3.7. Although the plot is not as complete as we had hoped, lacking significantly in the plateau region, solubility of the starting material precluded the collection of more data points. However, the plot does fulfil our requirement for an estimation of the phosphoramidate's stability under transcription conditions, as well as its differing rates of hydrolysis in high and low pH solutions

PH	k_{obs} (s^{-1})	Log k_{obs} (s^{-1})	$t_{1/2}$ (h)
3	4.55E-04	-3.3	0.42
3.5	2.40E-04	-3.6	0.80
7.2	1.11E-04	-4.0	1.7
8	7.71E-05	-4.1	2.5
9	3.21E-05	-4.5	6
9.8	3.25E-06	-5.5	59
10.5	2.10E-06	-5.7	92

Table:3.3 Data used to plot Fig:3.4 and the corresponding $t_{1/2}$ values.

The pH rate profile shows a classic alkyl phosphoramidate hydrolysis pattern, as can be seen through comparison to earlier examples in Chapter 2. The shape of the plot can be explained by the ionic species in prevalence at each pH value. An initially slow reaction at high pH displays an elevation in rate as pH is lowered, corresponding to the formation of the more reactive monoanionic GANP species. At approximately pH 7 the GANP species present is uniformly the monoanion, consequentially the rate of reaction plateaus. Formation of the more reactive neutral species at lower pH results in a corresponding increase in the observed rate of reaction. We were able to gain results to define both the low and high pH sections of the plot, however for the solubility issues detailed above we were unable to acquire data to fully delineate the plateau region.

3.9 Results and discussion

The half-life of the phosphoramidate hydrolysis was found to range from 0.4 hours at pH 3, to a much slower 59 hours (~2.5 days) at pH 10.5. The most significant result for our research was the 2.5 h half-life of the reaction at pH 8. As transcription reactions rarely have duration over 2.5 hours we were confident that the phosphoramidate was stable enough for GANP to be able to initiate RNA polymerisation. In addition the short half-life at low pH indicates our strategy for the efficient removal of the phosphate-masking group *via* pH control has a good chance of success. Following the success of this investigation we were encouraged to carry out transcription studies with GANP to assess its ability to initiate polymerisation reactions.

4.0 Transcription reactions and product analysis

4.1 Chapter contents

As stated in the Chapter 3 (Section 3.9) the half-life of GANP hydrolysis is 2.5 h under transcription reaction conditions. As we intended to incubate the transcription reactions for 2 h, we felt the stability of GANP fit well into the 'window of opportunity' for its use as an initiator (Section 2.23). Initially we wanted to determine whether the novel nucleotide would have any activity. Upon confirming its functionality we then extended our study to assess the optimum concentration of GANP to use in order to maximise phosphoramidate incorporation. The details of our studies using GANP can be found in a later chapter (Chapter 5), what follows is an account of the transcription reaction conditions we used in all our investigations. This chapter also describes the analytical techniques we employed to determine the total RNA yield of each transcription reaction, as well as the bioconjugative chemistry and subsequent PAGE assay used to determine the percentage of RNA initiated by GANP, Amino–G or Azido–G.

4.2 Transcription reactions

Transcription reactions are used for the synthesis of RNA chains, also known as transcripts, of a predetermined sequence. Initially discovered as an *in vivo* process, it has since been demonstrated that the reaction can be carried out *in vitro*[3,6]. It is the more convenient and adaptable *in vitro* methodology that we will be employing in our research. In both *in vivo* and *in vitro* protocols the reaction is enzyme mediated and utilizes a double stranded DNA template for the polymerisation of nucleotide triphosphates into RNA transcripts.

In the synthesis of 5′-triphosphate terminated RNA transcripts the only nucleotides present in the transcription reaction mixture (aside from the DNA template) are guanosine, adenosine, uridine, and cytosine 5′-triphosphates (GTP, ATP, UTP, CTP). These triphosphates are incorporated into both the terminus and body of the RNA chains

to form the transcript products. The DNA template we used contained a T7 promoter region that increases the total RNA yield of the reaction, but has the requirement that the polymerisation reaction is initiated by a guanosine derivative. Therefore most RNA transcripts produced using T7 RNAP have a guanine base at their 5′–terminus.

4.3 Novel nucleotides initiating transcription reactions

The T7 RNA polymerase enzyme, chosen owing to its proven ability to initiate RNA polymerisation with novel nucleotide compounds, mediates the transcription reactions we employed. The source of this flexibility lies in the nature of the enzyme's active site, specifically in the tolerance shown for modifications at the 5′ positions of nucleotides. A more in-depth discussion on the enzyme and other nucleotides included in the reaction mixture can be found in Chapter 1 (Section 1.6 onwards).

It has been shown that numerous novel-initiating nucleotides can be used to synthesise RNA incorporating a novel group at the 5′–terminus (Section 1.26). The novel initiator is added as part of the reaction mixture and competes with GTP to initiate polymerisation reactions. Observing the percentage of those RNA chains produced which possess the novel functional group allows the researcher to assess how successfully the novel compound competes.

4.4 Transcription reaction conditions

Transcription reactions were carried out under standard T7 RNA polymerase conditions as defined by Promega[60]:

40 mM Tris (pH 7.9)*,

6 mM $MgCl_2$*,

10 mM NaCl*,

2 mM spermidine*,

10 mM DTT,

350-500 nM dsDNA template,

1.25 mM of each NTP,

X mM Novel nucleotide**,

420 nM T7 RNA Polymerase***,

*prepared as a 5 × T7 Buffer stock.

**addition of a novel nucleotide concentration varied as required. 0 mM in standard reactions.

***over expressed *in vivo* and purified *via* a literature method[61].

The reactions were incubated at 37 °C for 2.5 h. After this period the reaction was halted by DNA template degradation with the DNA specific nuclease DNase (1Unit) and further incubation at 37 °C for 30 mins. The various purification steps detailed below were then employed to isolate the RNA chains produced.

The DNA template that we used encodes for the acyl-transferase ribozyme $ATRib^{TL}$, a 75 mer RNA sequence previously developed and studied by Suga *et al*[45]. During their research they tried with limited success to use Amino–G to initiate T7 RNAP mediated transcription reactions, thereby incorporating an amine group at the 5′–terminus of the ribozyme. We chose to use the $ATRib^{TL}$ ribozyme sequence for our research as this would allow us to compare our results for the level of incorporation achieved to those previously reported by Suga *et al.* The DNA template was synthesised *via* a PCR reaction using chemically synthesised template and primer oligonucleotides.

4.5 Product purification

As illustrated above transcription reactions contain a wide variety of reagents, also an array of by-products are formed along with the desired RNA transcripts. To allow us to carry out further studies on the RNA products it was necessary for us to first isolate them from the crude reaction mixture. To achieve this we used the standard oligonucleotide purification method of denaturing polyacrylamide gel electrophoresis (PAGE), using a tris borate buffer (pH 8). This technique uses a polyacrylamide gel as an electrophoresis support, enabling the separation of nucleic acids and other reaction components as functions of their size and charge. The amount of cross-linking between polymer chains in the gel can be varied according to the length of the transcripts that require separation.

Larger nucleic acid chains can be separated on gels with a higher percentage of cross-linking, while the separation of smaller chains requires a lower percentage.

Following electrophoresis the nucleic acid bands were visualised by UV shadowing as described in the figure below.

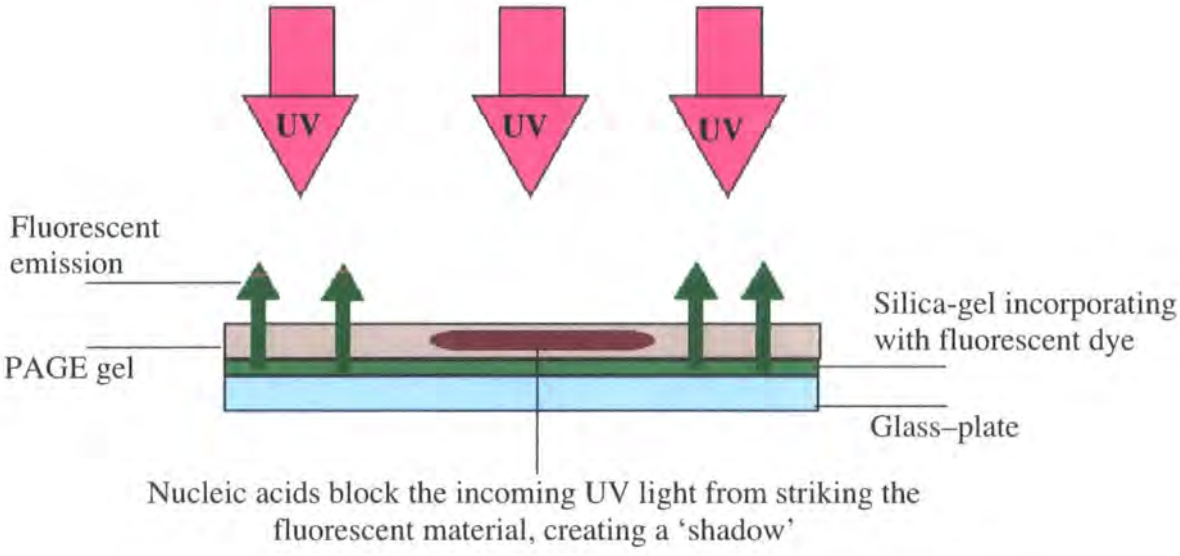


Fig:4.1 Diagram of UV shadowing technique used to locate UV active materials bound in a PAGE gel.

The areas of the gel containing our desired RNA products were then excised from the gel and macerated. The RNA was then eluted from the gel fragments into $NaCl_{(aq)}$ (300 mM), through stirring overnight on an end–over–end rotator. Removal of the gel fragments from the solution was achieved using spin–columns fitted with porous membranes. The eluent from these columns was a very dilute solution of RNA. Before we could carry out further studies on the RNA products we first had to concentrate the solution, which was carried out by first precipitating the transcripts through the addition of ethanol (2.2 volumes) and cooling the resulting solution to –20 °C. The solid that formed was collected by centrifugation, whereupon it was washed with a small volume of 70 % ethanol to remove excess salt before air-drying then re-dissolving the transcripts in a small volume of DEPC water. The resulting RNA stock was then available for use in bioconjugation reactions enabling us to assess the percentage incorporation of novel nucleotide.

4.6 RNA visualisation and analysis

To investigate the effect GANP had on the products of transcription reactions it was necessary that we could observe and analyze the RNA being studied. Although RNA is UV active, the small quantities of material used in a bioconjugation assay were too low to detect through convenient UV techniques, making modification of the RNA transcripts necessary. Rather than chemically modify the RNA through a potentially inefficient methodology we decided to use a radioactive isotope of phosphorus to label the reaction products.

Although the use of radio–isotopes does pose both technical and health and safety considerations, we felt that the sensitivity and reliability of the technique, coupled with the low levels of exposure necessary to gain useful results, outweighed the risks. We chose to body label the RNA during synthesis using α-^{32}P UTP as this method yields high sensitivity while not being labour intensive or time consuming. Also owing to the labelling taking place during the transcription reaction the time spent handling highly active radiation sources is reduced to a minimum.

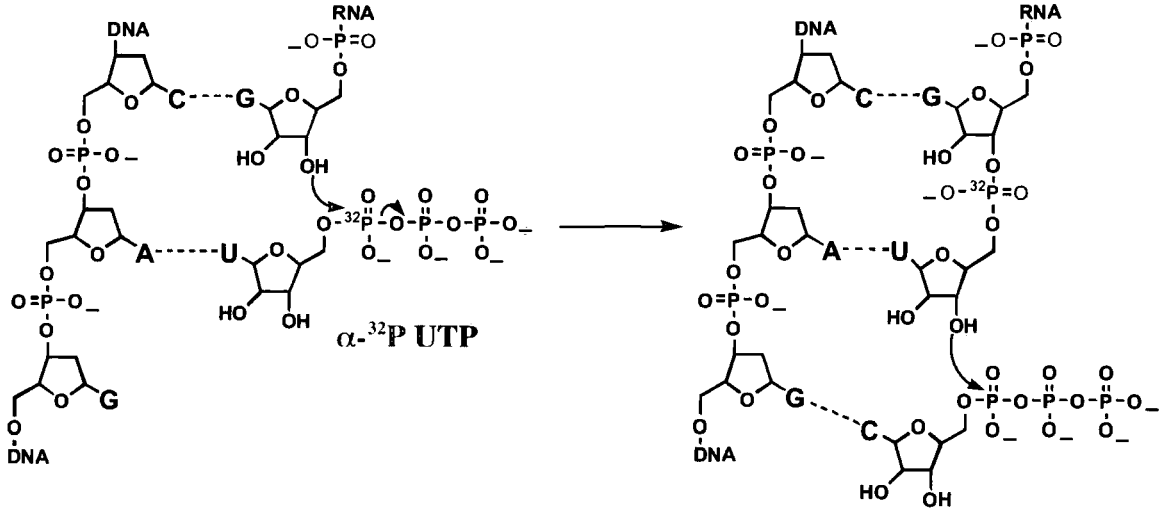


Fig:4.2 Proposed mechanism for the incorporation of α^{32}P labelled UTP into the body of RNA transcripts as they are synthesised.

The ^{32}P isotope undergoes β-decay, emitting an electron and a proton as a neutron degrades, forming a ^{32}S atom. The half-life of the decay is 14.3 days, providing ample time to carry out labelled transcription reactions, then bioconjugate and study the RNA products. Emission is of low range and moderate energy; hence all experimental work was carried out behind Perspex screens and work areas were monitored for residual radioactivity with a Gieger-counter. Radioisotope containing bands in PAGE gels were visualised *via* Fujifilm phosphor imaging plates.

4.7 Phosphor imaging plates[62]

Fluorescent substances are those compounds that emit light upon a specific stimulation. The stimuli can come from radiation, UV light, heat, a mechanical force or chemical reaction. A group of these substances, often referred to as phosphors, are commonly used as powders and have practical applications. The light emitting behaviour of phosphors can be categorised into a number of phenomena. If the emission of light ceases immediately upon removal of the stimuli, the phenomenon is known as 'fluorescence'. In some cases light emission continues for a short while after the stimulation ceases. This phenomenon is known as 'phosphorescence'. 'Luminescence' incorporates both of these light-emitting behaviours.

Which characteristic behaviour is observed can be accurately adjusted through alteration of the phosphor composition and manufacturing process. The phosphor used in the Imaging Plates used to visualise our radioactive samples in PAGE gels utilizes the 'photostimulated luminescence' (PSL) phenomenon[62]. The discovery of this phenomenon has been attributed to Becquerel in the mid-19th century and is neither fluorescent nor phosphorescent. The PSL phenomenon can be attributed to a substance that upon stimulation by, for example radiation, exhibit luminescence; then upon a second stimulation by light of a longer wavelength than the initial luminescence the phosphor emits light for a second time.

The photo–sensitive phosphor used in a Fujifilm Imaging Plate is composed of barium fluorobromide containing trace amounts of bivalent europium acting as a luminescence

centre[62]. The plate is made flexible, and therefore durable, by manufacturing the phosphor as very small crystals (BaFBr: Eu^{2+}, grain size ~5 μm) and uniformly coating it on to a polyester support film. A thin polymer coating ensures that the phosphor is protected.

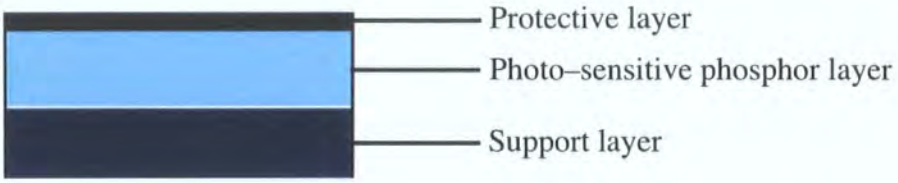


Fig:4.3 Cross-section showing layers of material present in a typical Phosphor Imaging Plate[62].

Through the use of this technology we can stimulate the phosphor compound in the IP plate with radioactive RNA samples still held in a PAGE gel. Exposure of sample to the Imaging Plate is performed in a manner similar to conventional photographic film; the exposed Plate is then scanned with a laser beam of red light while being conveyed with high accuracy in a phosphor reader. Collecting and recording the resultant emission of light forms an image of the gel and the labelled material contained within it. In addition, as the strength of the initial radioactive stimulation determines the intensity of the final light emission, the relative amounts of RNA in each observed band can be assessed.

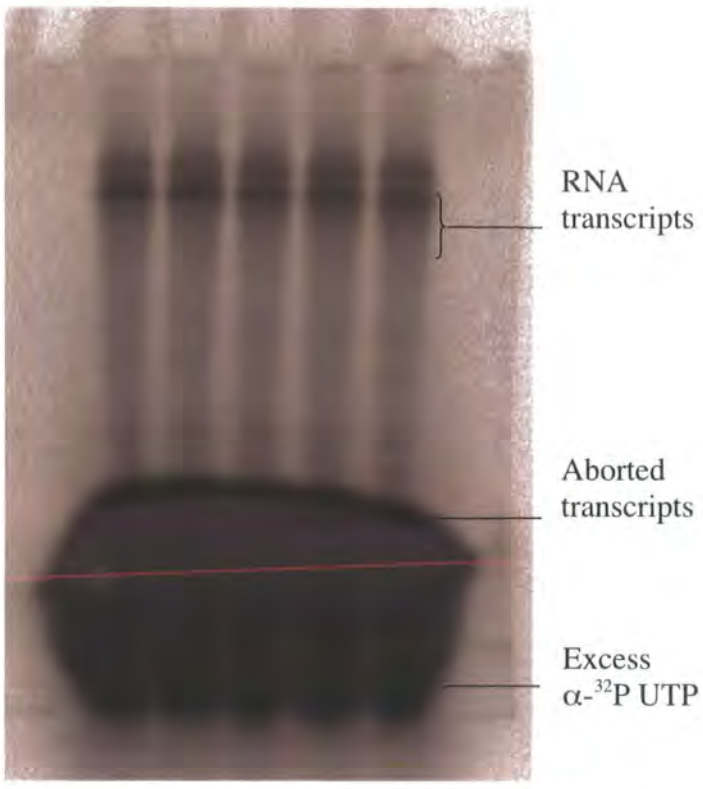


Fig:4.4 A typical image of ^{32}P labelled RNA still bound within a PAGE gel, as collected by a Phosphor Image Plate Reader. The image clearly shows all radioactive species present, including the product RNA, short length RNA by-products, and unincorporated label.

The image above (Fig:4.4) shows a phosphor plate after exposure to a PAGE gel that had been used to purify ^{32}P labelled RNA transcripts from by-products and reagents. The more intensely black regions are those areas of the gel that contained more ^{32}P label and hence resulted in greater stimulation of the phosphor compound. Obviously the regions we were most interested in were the bands relating to the labelled ATRib^{TL} transcripts. The diffuse, high intensity area at the bottom of the gel is the result of residual, highly mobile, α-^{32}P UTP in the transcription reaction. A faint outline of the polyacrylamide gel, including loading wells, is also visible.

4.8 Assessment of novel nucleotide incorporation levels.

To assess the effectiveness of our novel nucleotides at initiating RNA synthesis we needed to compare the relative amounts of RNA transcripts primed by the nucleotide being studied versus those primed with GTP. To make this evaluation we required a

method of distinguishing between these two products of the transcription reaction. As the novel nucleotide primed transcripts possess a reactive 5′–amino group to which a variety of reporter groups or affinity tags could be bioconjugated, and the GTP primed transcripts do not, a number of routes to differentiate between the two products of the transcription reaction were open to us. The method we chose to employ, owing to its reliability and ease of use, was the bioconjugation of a biotin-containing group to the novel nucleotide primed transcripts *via* the terminal amine.

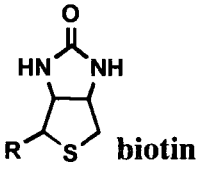


Fig:4.5 Although a small molecule biotin forms very strong interactions with the proteins avidin and streptavidin. Owing to this interaction Biotin is a commonly used affinity tag for biological molecules.

4.10 Biotin affinity tags

Biotin is an affinity tag that takes part in an extremely strong non-covalent binding interaction with streptavidin, a 60 kDa protein, to form a conjugate of greatly increased mass. The dissociation constant of the biotin-streptavidin conjugate (~1.3×10^{-15} M) is very large; significantly the coupling is too strong to be broken by gel electrophoresis. Therefore the increase in mass resulting from streptavidin conjugation to a biotinylated RNA molecule can be exploited to separate otherwise similar biotinylated and non-biotinylated transcripts *via* PAGE. In our studies biotinylation was carried out using an amine specific bioconjugation reaction, thus ensuring only those transcripts initiated with GANP/Amino–G/Azido–G and subsequently biotinylated could form conjugates when introduced to streptavidin. Streptavidin is introduced to the transcript mixture after the biotinylation reaction has taken place but prior to gel loading. Upon electrophoresis the biotinylated RNA–streptavidin conjugate exhibits a reduced rate of migration compared to the unmodified RNA as a consequence of its greatly increased mass.

The amine specific biotinylation reagent we used was 3-sulfo-NHS biotin:

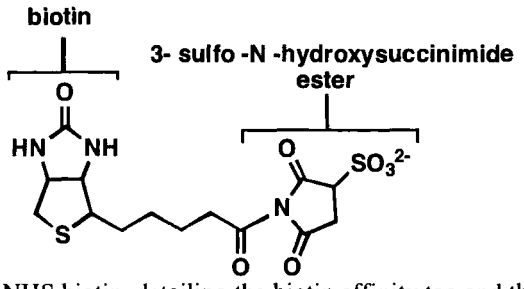


Fig:4.6 3-sulfo-NHS biotin, detailing the biotin affinity tag and the succinimide ester that confers the specificity of the reagent toward reaction with amines.

Although biotin-streptavidin binding is extremely strong, the proteins binding site does not lie at the surface, a consequence of this is that larger biotinylated compounds may bind inefficiently owing to steric constraints, leading to inaccurate results. It has therefore become common to design biotinylation agents with a short spacer arm between the reactive-group. We chose not to employ such a method as we determined the flexibility of RNA transcripts and the position of the tag on it (*i.e.* at the terminus) to be ample to provide efficient binding to the protein active site.

The presence of the anionic sulfate on the succinimide sub–structure is necessary to ensure the otherwise hydrophobic biotinylating agent is water-soluble. As the bioconjugation reaction mechanism is one of nucleophilic substitution, carrying out the biotinylation in water raises the issue of selectivity, to counter this the reaction mixture was buffered to pH 8 using a non-nucleophilic HEPES buffer, ensuring a significant proportion of the amine was in its reactive deprotonated form.

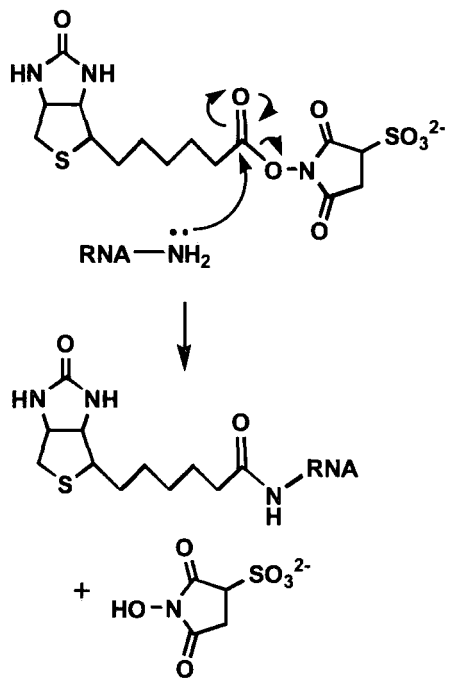


Fig:4.7 Proposed mechanism for the bioconjugation of the biotin affinity tag to 5'-AminoRNA. Through the use of this reaction we can selectively tag those RNA transcripts primed by GANP/ Amino–G/ Azido–G. Hence in conjunction with streptavidin we can run a PAGE assay to qualify the nucleotides level of incorporation.

As the amounts of RNA used in each bioconjugation reaction were so small we had no way of assessing the yield of the biotinylation reaction. Hence, biotinylation reagent was used in excess to ensure the yield of reaction was as close as possible to the assumed value of 100 %.

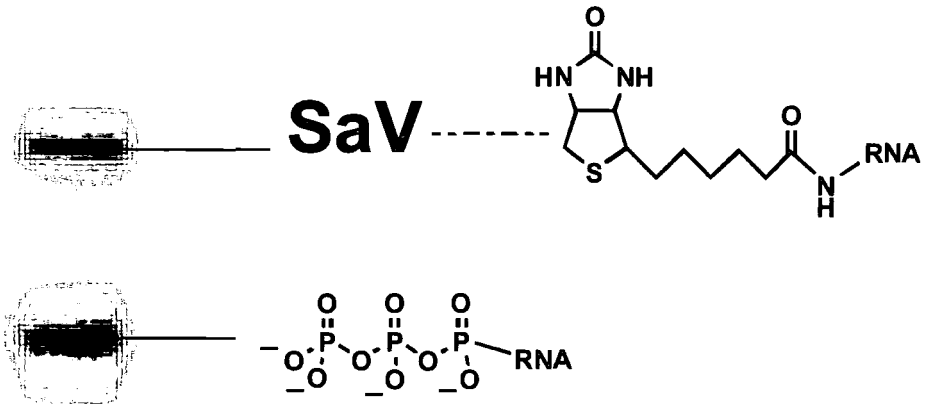


Fig: 4.8 Portion of PAGE gel showing separation of two radio labelled RNA transcript populations. Comparrison of the intensities of the two bands allows the level of novel nucleotide incorporation in the RNA population.

The IP plate image above shows the separation of RNA chains achieved through the use of a biotin affinity tag in conjunction with streptavidin. By analyzing the purification gels, it is possible to quantify the relative intensities of the RNA bands. Therefore we can assess qualitatively what percentage of the transcription product has been initiated by a novel–nucleotide.

The methods and techniques detailed in this chapter were used in the majority of our research using transcription reactions. The following chapters provide a more specific discussion of the research undertaken and results collected using GANP, Amino–G, and Azido–G novel nucleotides.

5.0 GANP transcription Study

5.1 Research targets

The purpose of this chapter is to detail our study into the ability of GANP to initiate transcription reactions and in doing so produce 5′–AminoRNA. We began by undertaking transcription reactions containing a crude sample of GANP *via* the methodology set out in the previous chapter (Section 4.2). The source of the crude GANP sample was the aqueous phosphorylation reaction described in Chapter 2, specifically it was the solid produced after the reaction mixture had the solvent removed under reduced pressure. During these preliminary reactions we faced our first obstacle in the form of residual novel nucleotide in the reaction mixture hindering the RNA purification process. Through the use of an additional gel filtration step we were able to overcome this problem and develop a standard methodology for the use of GANP that applied to use of both crude and purified GANP. The total RNA yields and levels of incorporation observed when using the crude sample indicated the use of purified GANP would be more successful. This was not the case however, as we observed that the results observed showed a significant dependency on the purity of GANP used.

The targets for this section of our study were:

- To establish whether GANP can be used to prime T7 RNA polymerase mediated transcription reactions, resulting in the synthesis of 5′-AminoRNA.
- If successful, can this reaction be optimised to maximise 5′–AminoRNA production?

Our primary aim was to ascertain whether or not GANP could initiate transcription reactions, and thus synthesise 5′–amino terminated RNA. Once the activity was confirmed we then wanted to optimise the conditions for GANP's use. The criteria for the optimisation are dependent on the final use for the RNA population. If a large

quantity of 5′-AminoRNA is required, regardless of the overall level of incorporation, the conditions used may be very different from those used to synthesise an RNA population for another purpose, *e.g.* requiring a high level of incorporation regardless of yield. By analyzing both the total RNA yield and the level of GANP incorporation resulting from each reaction we intended to be able to specify conditions required for both extremes of GANP use. Our benchmark for success would be the levels of incorporation/ total RNA yield observed through the use of close analogues to GANP such as GMPS, GSMP, and other 5′-amino nucleotides (see Section 1.26–35).

5.2 Hypothesised behaviour of GANP

We hypothesised that GANP would behave in a similar manner to other GMP analogues such as GMPS, GSMP, *i.e.*:

1. At low concentrations of GANP an increase in chain initiation caused by the presence of heightened levels of guanosine moieties will result in a rise in total RNA yield.
2. As novel nucleotide concentration is increased relative to NTP concentration, total RNA yield will be reduced as competitive binding in the chain elongation stage outweighs the effect noted in point 1.
3. Percentage of chains initiated by the novel nucleotide will increase in proportion to the concentration of GANP used.
4. An optimum will be found where a compromise between reduction in total RNA yield and the percentage of GANP incorporation can be made.

Phosphorothioate analogues of GANP are quoted as being capable of being incorporated in 60–95 % of RNA chains with a cost of lowering total RNA yield as the level of incorporation is increased. As GANP possesses a phosphoamidate group with a half-life approximately equal to the 2.5 h reaction time under transcription conditions (Section 3.9), we expected the observed levels of incorporation to be comparable to those of GMPS, GSMP.

5.3 Competition with GTP

To counter the competition between GANP and GTP to initiate transcription reactions, the phosphoramidate was used in a higher ratio, typically 1.25 mM GTP: >2.5 mM GANP. By swaying the competition in favour of GANP we expected to increase the subsequent yield of 5'-AminoRNA. Optimization of the GANP concentration to factor in other competing effects could be achieved through a systematic investigation once the methodology to determine total RNA yield and percentages of incorporation were in place and proven to be accurate and robust.

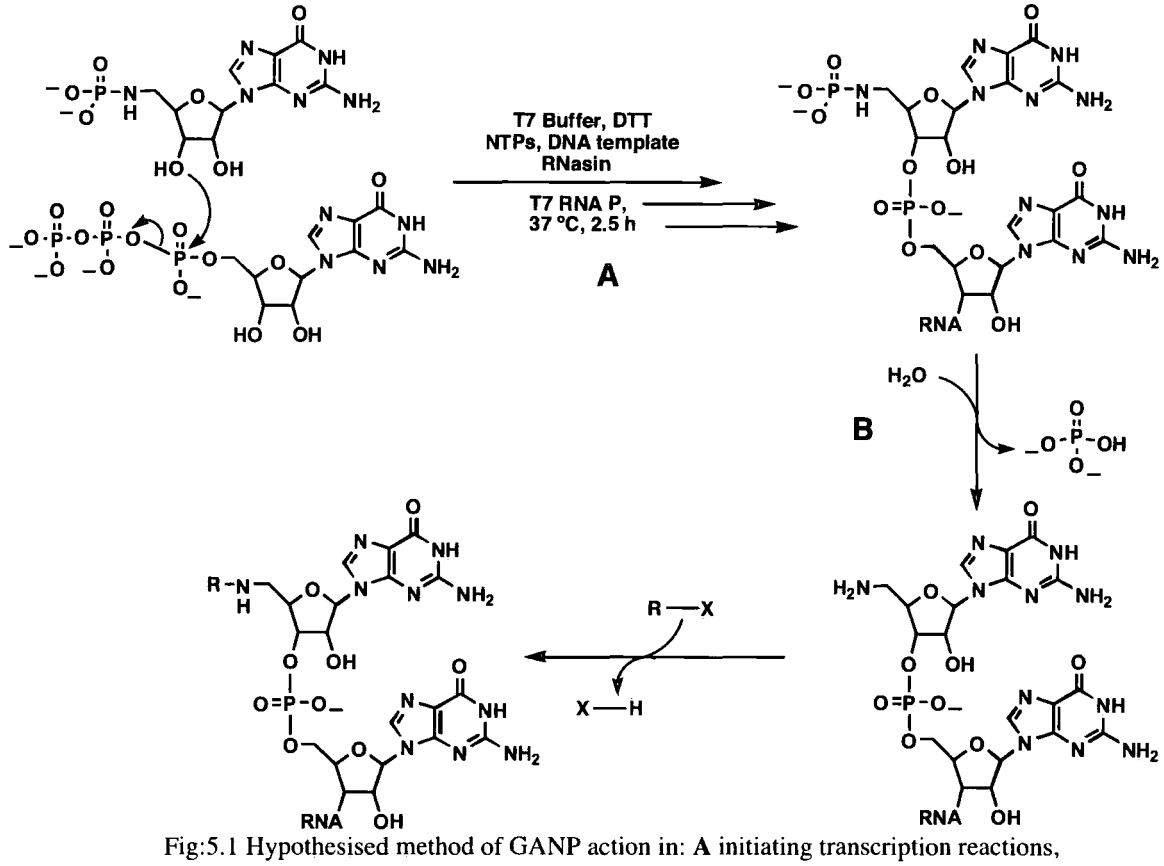


Fig:5.1 Hypothesised method of GANP action in: **A** initiating transcription reactions, and **B** hydrolysing to reveal a reactive terminal amine group.

5.4 First attempt at use of GANP

Unfortunately owing to the unavailability of analytically pure GANP (see Section 2.54) we began our transcription reaction study using a crude sample. The sample used was the redissolved crude reaction mixture from an aqueous phosphorylation reaction.

Phosphorus NMR indicated 90% of the $POCl_3$ starting material had successfully been converted to the phosphoramidate (see Fig5.2). Other impurities present included unreacted Amino–G, phosphate, NaOH, and NaCl. Of these only the Amino–G is suspected to have a significant effect on the transcription reaction, as will be discussed at a later point.

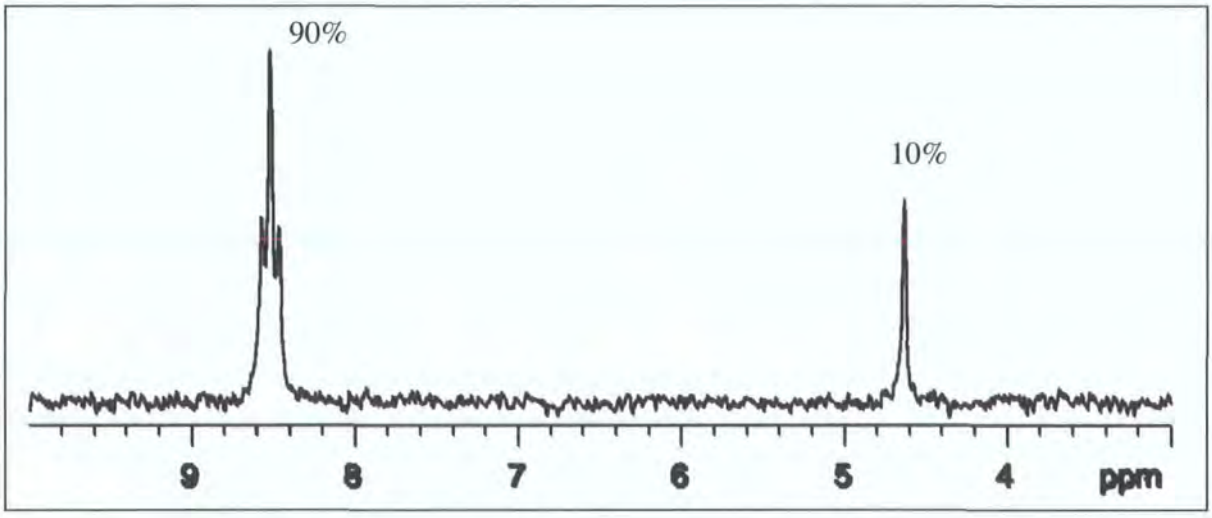


Fig:5.2 ^{31}P NMR spectra of crude GANP in D_2O featuring signals related to two phosphorus containing species: the GANP phosphoramidate triplet at ~8.5 ppm, and phosphate at ~4.7 ppm in a ratio of 9:1[57,59].

As the amount of each impurity in the sample was unknown we could not determine the exact amount of the novel nucleotide we were adding to the reaction mixture. Therefore when determining the concentration of our stock GANP solution, and hence the amount of phosphoramidate in each subsequent reaction mixture, we assumed the crude product to be 100 % GANP. This is obviously unrealistic but allowed us to gain valuable insight into the use of GANP until a cleaner source became available.

Our first transcription experiment consisted of two reactions, a control with no GANP present, and a transcription containing 10mM GANP. The other constituents of the reaction mixture were those laid out in Section 4.2. To ensure the reactions were comparable the level of UTP*, NTPs, DTT etc present in each transcription must be constant, this was achieved by using a stock of all reaction components other than GANP

and the T7 RNA polymerase enzyme. This stock was then distributed between two reaction vessels, one containing 10 mM GANP, the other without GANP. The transcription reactions were initiated through the addition of T7 RNAP, the mixture was then incubated at 37 °C for 2.5 h. Following incubation, during which the transcription reaction took place (in those cases where the reaction was successful) and DNA degradation was performed, the reaction mixtures were loaded directly onto an 8% cross-linked acrylamide gel and electrophoresed for 1.5 h at 30 W. Unfortunately, as the Fig:5.2 shows the reaction was not as successful as we had hoped.

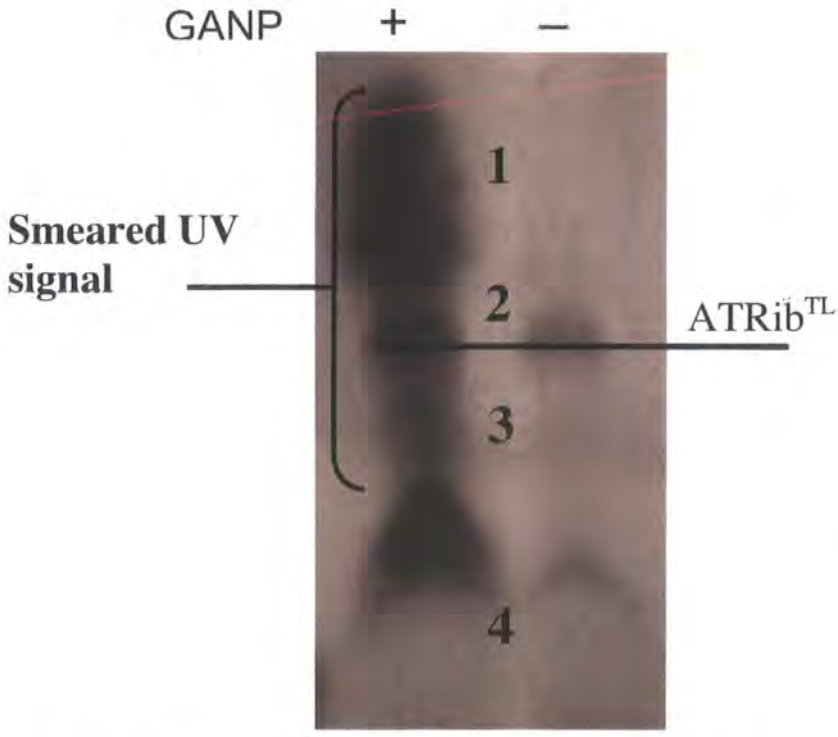


Fig:5.2 A UV shadowing image of the 8% PAGE gel used to purify initial GANP containing transcription attempts and control transcription reactions. The large smear present in the +GANP lane obscures clear visualisation of the ATRib^{TL} band, before further studies could be carried out the source the substance responsible had to be determined.

Although both + and - GANP lanes contain a band of what was likely to be ATRib^{TL} (band 2) transcripts, a significant difference also exists, specifically, the + GANP lane contains a smear of UV active material (see Fig:5.2, band 1+3). The shadow of this substance begins in the abortive transcript region (band 3) and forming a diffuse band of

considerable size above the full transcript signal (band 1). As the two reaction mixtures were identical other than the inclusion of crude GANP it is logical to propose the smear originates from the novel nucleotide sample.

5.5 Origin of UV active substance

We would expect that GANP, as an anionic nucleotide species, would migrate only slightly slower than the excess NTPs present in the reaction mixture. It is evident from the large area covered and the evident range of migration speeds covered by the UV smear (bands 1+3) that GANP was not solely responsible. Species migrating to a lesser degree than the RNA transcripts would have to be either larger, possibly a dimer of the intended transcript, or posses a lower negative charge. As the RNA in the reaction mixture is denatured through heating to 95 °C in 4 M urea loading buffer prior to gel loading, the presence of dimers is unlikely. Therefore we formed another hypothesis for source of the unwanted UV material

We know the anionic phosphoramidate group is hydrolytically labile with a half-life of ~2.5 h under transcription reaction conditions, hence after the reaction has been completed a significant proportion of the GANP present would have hydrolysed to form the corresponding amine. Without the dianionic phosphate group, migration of the amino-nucleoside would be much slower than that of GANP. Hence, it is possible that the presence of both Amino-G and GANP in the reaction mixture undergoing electrophoresis would explain the observed UV traces above the $ATRib^{TL}$ band (band 1). We hypothesised that residual GANP in the reaction mixture would give rise to the lower UV trace (band 3), while hydrolysis of the phosphoramidate on the gel would explain the smear of the signal. By running transcription reactions containing a range of GANP concentrations we hoped to observe a concurrent increase in the unidentified signals intensity, thereby confirming the source of the UV material as GANP and Amino–G.

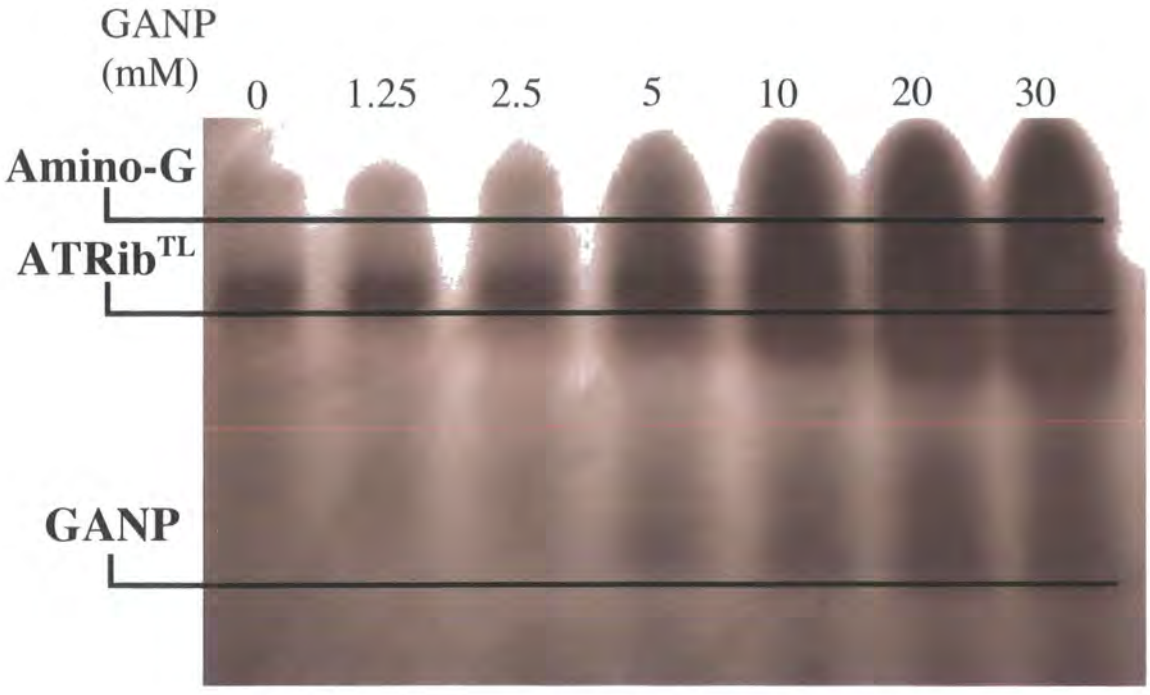


Fig:5.3 UV shadowing image showing PAGE separation of several reaction mixtures containing increasing levels of GANP.

As predicted an increase in the intensity of the diffuse signal, which obscures the ATRib^{TL} band, is observed as GANP concentration is increased. In the higher concentration reaction mixtures a significant problem arises where the ATRib^{TL} product band becomes obscured. This has a clear effect on the ease of band identification, as well as other problems that will be discussed later. To test our hypothesis that the signal obscuring the ATRib^{TL} band was partly due to Amino–G we compared the migration of Amino–G to that of a GANP containing transcription reaction. If the hypothesis proved correct we would observe a UV active substance in the Amino–G lane coincident with the upper most section of the smear present in the GANP containing lane.

GANP AMINE

Fig:5.4 UV shadowing image of PAGE separation of GANP containing transcription reaction, and a sample of Amino–G. As hypothesised the Amino–G smear is co-incident with the slowest migrating substance in the GANP lane.

The GANP lane in the image above (Fig:5.4) displays the same UV active smear observed previously, while the Amino–G lane featured a single diffuse band of low mobility through the gel. The comparability between the migration of the amine and part of the GANP signal gives support to our hypothesis. With no other candidate to hand we can confidently attribute the observed UV smear to GANP and the product of its hydrolysis, Amino–G.

5.6 Problems with residual GANP

The presence of the residual guanosine derivatives in the gel causes a number of problems:

1. Identification of the ATRib^{TL} band for excision is made more difficult.

2. Excised gel will contain Amino-G as well as ATRib^{TL} transcripts.

In most cases the ATRib^{TL} transcript band can be seen regardless of interference from GANP and Amino–G signals, therefore the most significant issue we had to face was the presence of excess amine in our excised RNA samples. This could be a major problem as our method of analysing the efficiency of GANP incorporation was *via* an amine-specific biotinylation reaction and Amino-G also contains a reactive amine. If our RNA transcript samples contained an unknown quantity of Amino–G, the results gained from any biotinylation experiments would be invalid as the amount of biotinylation reagent available for reaction with the 5′–AminoRNA would not be constant across the experiments carried out.

5.7 Removal of residual novel mononucleotide and nucleoside

The most obvious method for separating the ATRib^{TL} transcripts from the residual GANP/ Amino–G was to run the electrophoresis for a longer period to increase the distance between each signal on the gel. However because the impurities migrate over such a wide area of the gel it was not convenient to fully separate the signals.

As the novel nucleotide impurities are very much smaller than our RNA product we believed gel filtration of the reaction mixture could remove much of the unwanted GANP and Amino–G before the PAGE step. Owing to the small volume of the reaction mixtures, and as we were using isotopic labelling, it was not practicable to use a gel filtration chromatography column of large capacity such as those used during the synthesis of GANP (see Section 2.57). Instead we utilized 0.6 mL Zeba Desalt spin columns purchased from Pierce to carry out the filtration. These columns are prepacked with the Zeba gel filtration support and once loaded with the reaction mixture use a centrifuge to force the eluent through the column. These columns had the added benefit of being disposable, an important factor as the residual UTP* would also be retained in the gel filtration matrix, precluding their reusability.

As with the Sephadex size exclusion media used previously (see Section 2.57), Zeba is a porous material that retards the elution of small molecules. Hence, the bulk of

mononucleotides present are retained in the spin column while RNA transcripts and larger molecules are eluted. This step was undertaken between halting the reaction with RNase free DNase enzyme and diluting the reaction mixture with urea loading buffer, the size of columns used was ideal for our needs, and provided a fast and efficient solution to our problems by greatly reducing the amount of residual GANP and Amino–G present, as can be seen in Fig:5.5.

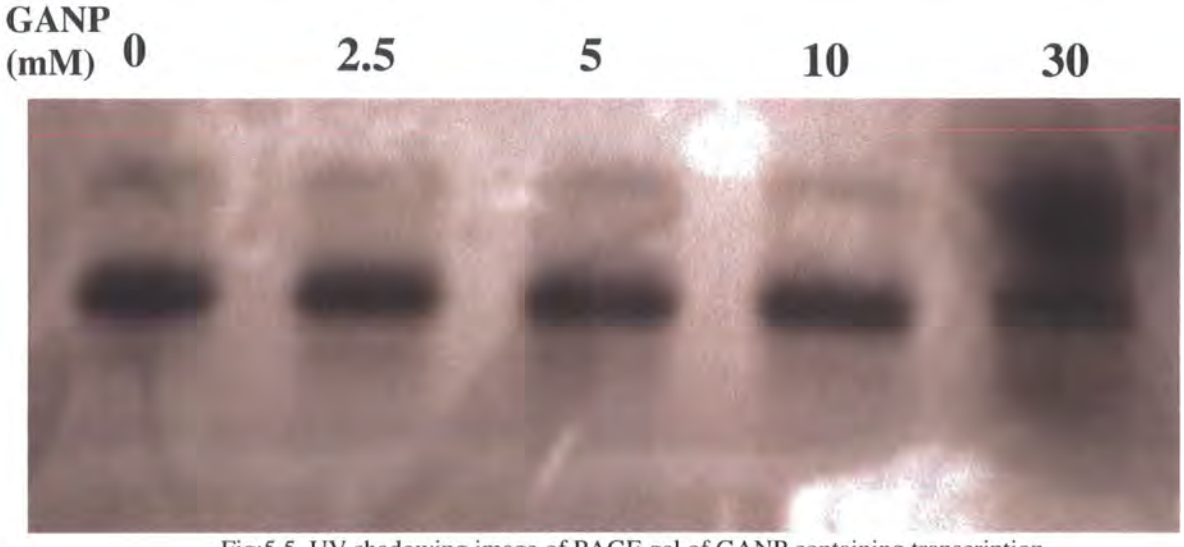


Fig:5.5 UV shadowing image of PAGE gel of GANP containing transcription reaction mixtures after filtration of the crude mixture through Zeba Desalt columns. The lanes are clearer than previous GANP containing images owing to the removal of residual small molecules prior to gel loading by gel filtration spin columns.

Significant removal of residual novel nucleotides was achieved *via* this method in most cases, only when there was a very high concentration of GANP present (30 mM) were novel nucleotides observed. Elution of small molecules from the column occurs when the medium is overloaded, complete removal could have been achieved by running the sample through another desalt column, but we felt this was unnecessary. With a method in place for the removal of the excess nucleotide we had developed a method for the purification of the synthesised RNA transcripts. Hence, we were now in a position to carry out biotinylation reactions and determine whether GANP was successfully initiating transcription reactions.

5.8 Selective biotinylation and PAGE assay

Following removal of excess Amino–G and GANP from the RNA transcripts by spin column, PAGE and precipitation (as described in Chapter 4), we possessed samples suitably clean for use in a bioconjugation reaction. In our initial strategy for use of GANP we had intended to incubate the GANP primed RNA transcripts in a solution of low pH to ensure complete removal of the phosphate masking group, as shown in Fig:5.3, however as we progressed with our studies this step became superfluous. The reason for this was the length of time the GANP–primed transcripts spent at high temperatures ~90 °C) and moderate pH during PAGE electrophoresis. At such high temperature it can be expected that the hydrolysis of the phosphoramidate group will occur with a half–life much lower than the 2.5 h (at pH 8) determined during the hydrolysis study detailed in Chapter 3 (Section 3.9). The degradation of GANP on a PAGE gel can be observed in the UV images provided above (Fig:5.3.) where the bands assigned to residual GANP are far less intense than those of it's hydrolysis product, Amino–G.

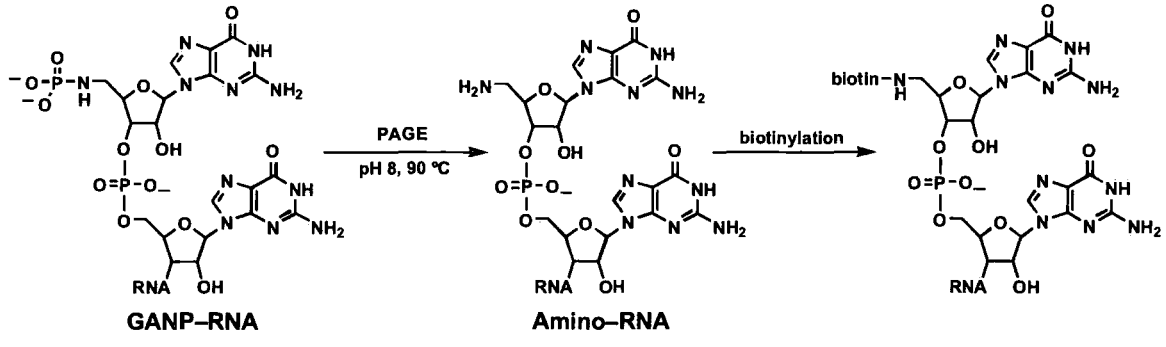


Fig:5.6 'Unmasking' of the terminal amine by hydrolysis of the phosphoramidate of GANP primed RNA during PAGE purification of the transcription reaction. Following which biotinylation of the AminoRNA can be carried out.

To confirm this hypothesis the products of a single transcription reaction containing GANP were split into two samples, one was frozen immediately while the other was incubated at 37 °C and pH 8 overnight ($t_{1/2}$ of hydrolysis = 2.5 h). The two samples were then purified by spin–column, PAGE and ethanol precipitation before biotinylation and streptavidin assay. If the PAGE (and other purification steps) resulted in an

inconsequential amount of hydrolysis we would expect to observe a higher level of GANP incorporation in the sample incubated overnight as more terminal amine would have been 'unmasked' and available for bioconjugation than in the 'quenched' sample. This was not the case however and the two samples showed almost identical levels of free amine, indicating the purification procedure had resulted in the maximal amount of phosphoramidate hydrolysis and that all the GANP primed RNA now possessed a reactive amine at their 5'–terminus.

To ensure no unwanted non-specific bioconjugation reactions were occurring, RNA synthesised under standard 0 mM GANP conditions were also subject to the biotinylation reaction, though no biotinylated product was ever observed. The amount of RNA used in each bioconjugation reaction was maintained constant across an experimental set by measuring the radioactivity of each sample. The amount of radiolabel per mole of RNA transcribed was an unknown quantity. However as the label was added to each transcription reaction in an experimental set came from a common stock and all RNA sequences isolated were theoretically identical in sequence, judging the relative amount of RNA in a sample by its radioactivity is reasonable. Obviously this is only true when comparing RNA produced at the same time, using the same stock solution of label and reagents, due to decay of the radio–label. As discussed previously the bioconjugation reaction used was the amine specific biotinylation of those RNA bands primed by GANP. The biotinylation reagent used was 3-sulfo-NHS biotin:

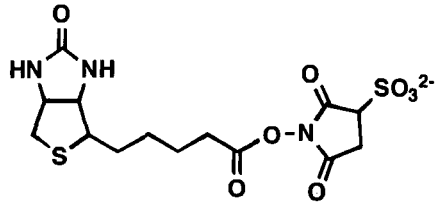


Fig:5.7 3-sulfo-NHS-biotin, the amine specific biotinylation reagent we utilized in our bioconjugation reactions.

The biotinylation reaction was carried out in a HEPES buffered aqueous solution (pH 8), incubated at 37 °C for 1 h. The biotinylating reagent was used in excess to ensure the maximum level of conjugation was achieved. By–products and residual reagents were

removed by precipitating the RNA transcripts in a manner similar to the ethanol precipitation detailed in Section 4.5. The mixtures of modified and non-modified transcripts were then redissolved in DEPC water before dilution with a urea and streptavidin loading-buffer. The transcripts were then separated by PAGE, after which the gel was dried, before the position and intensity of radioactive material in the gel was recorded through exposure to an IP plate. The resulting IP image is shown in Fig:5.7:

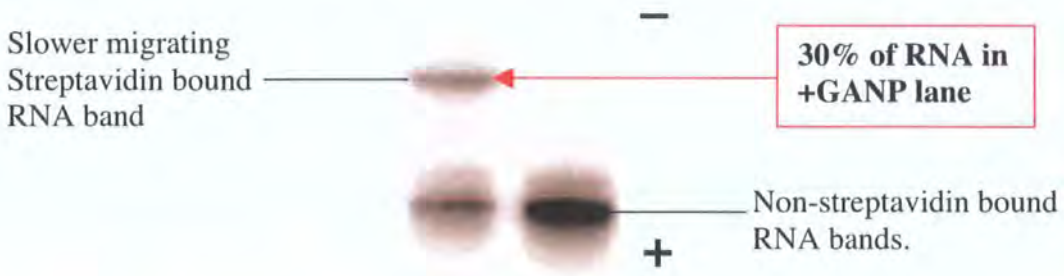


Fig:5.8 IP image of a PAGE gel showing the migration of two RNA samples, one transcribed in the presence of GANP, and one that was not. Both samples had been subjected to an amine specific biotinylation reaction and were electophoresed in the presence of streptavidin. The presence of two bands of differing migration speeds in the +GANP lane shows the biotinylation was successful in that sample and not the –GANP population.

The differing migration speeds of streptavidin bound and non-bound RNA in the +GANP lane causes separation of the two species, allowing an assessment of the percentage GANP incorporation that took place during the transcription reaction. In the case of the assay in Fig:5.8, 30% of the RNA in the +GANP lane binds to streptavidin results in the confirms that a two bands in the +GANP lane of differing migration speeds. The slower migrating band accounts for 30% of the total RNA yield in the +GANP lane (as assessed by determining the intensities of radioactive label in the two bands), which, assuming the biotinylation reaction modifies 100% of the AminoRNA in the sample, equates to 30% of the RNA sequences being initiated with GANP.

The faster migrating unbound RNA in both the +/–GANP lanes consist of those sequences whose synthesis was initiated by GTP. As they posses a triphosphate rather than an amine group at the 5′–terminus they are not biotinylated when incubated with the

NHS ester biotinylating reagent. Without a biotin tag attached they cannot bind to streptavidin and therefore migrate at a faster speed than the larger, conjugated species. The –GANP lane provides an important control assay as it illustrates that conjugation of an RNA molecule to streptavidin in this study can only occur through the biotinylation of RNA that have been initiated with GANP, and thus possess a terminal amine group. This is a crucial factor as all our assessments of the novel nucleotide's ability to incorporate into RNA sequences will be based on the fact that the slower migrating bands contain exclusively those RNA molecules initiated with GANP. The assay allowed us to realise our first target of demonstrating the ability of GANP to initiate transcription reactions, and thus cause the incorporation of a reactive amine at the 5′-terminus of the resulting RNA transcript. The initial level of incorporation of 30% is low by the standards of other novel nucleotides (see Table:1.3) but prior to optimisation is a promising value.

5.9 Crude GANP optimization

Research into methodologies similar to our own use of GANP, such as GMPS and GSMP incorporation, have found that ratios of between 8–4:1, novel nucleotide: GTP, provide optimal amounts of modified RNA. Owing to the structural and functional similarity between GANP and these other novel nucleotides we expected to find our optimum concentration within similar limits.

The first step in the optimisation process was to attempt to push the transcription reaction to failing point. As previously mentioned (Section:1.27) an excess amount of GANP present in the reaction was hypothesised to result in a reduced total RNA yield. It follows that a point will exist where the amount of RNA produced is so low as to be undetectable by both UV shadowing and isotopic labelling. By running transcription reactions containing high GANP concentrations we intended to discover where this point might lie.

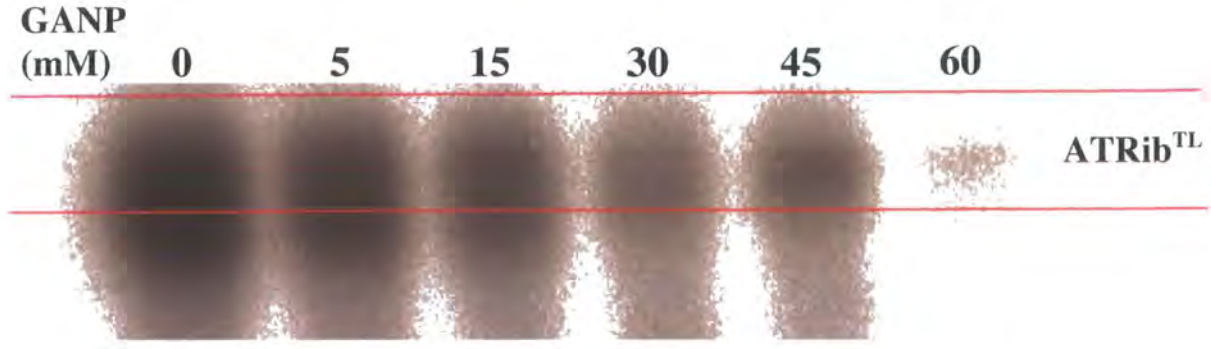


Fig:5.9 IP image of PAGE gel showing ATRib^{TL} products of transcription reactions. A low contrasting image was necessary to visualise the low RNA yield of the reactions containing higher concentrations of GANP.

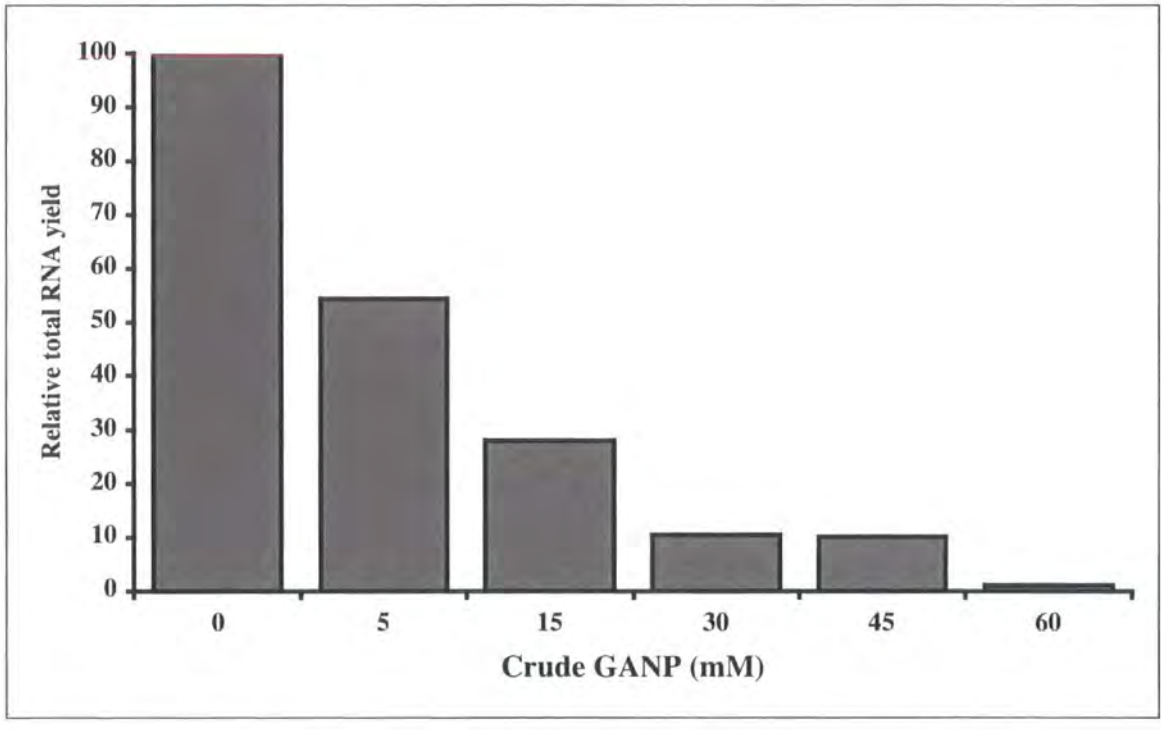


Fig:5.10 Plot of relative RNA yields for each concentration of GANP used. Yields assessed through intensity of ATRib^{TL} bands in IP plate image of PAGE gel (Fig:5.7). Yield of RNA from control (0 mM) reaction taken as 100% yield, all other intensities quoted relative to this value.

As expected, with increasing GANP concentrations total RNA yield is reduced, relative to the control reaction. Even when the reaction contains GANP at 60 mM some RNA transcripts are produced (1 % of control), though in such a small quantity that they would be very difficult to extract and use. Although we did not find a point where RNA

production is halted completely the experiment revealed that the optimum GANP concentration is likely to lie in the range 0–15 mM, hence subsequent investigations were carried out in and around this range.

5.10 Optimization experiments

Through the methods described above (Section 5.8) we were able to assess the changes in total RNA yield and level of GANP incorporation for the products of a number of transcription reactions each containing a different concentration of GANP. This enabled us to find the optimum amount of GANP necessary to maximise both the yield of RNA produced and the level of GANP incorporation. The IP images and results shown below are from one such experiment and are representative of data collected over the course of the optimization investigation. To ensure the reactions were comparable the level of radioactivity, NTPs, DTT, $MgCl_2$, NaCl, spermidine, Tris buffer, DNA template, and T7 RNAP present in each transcription must be constant, this was achieved by using a stock of all reaction components other than GANP and the T7 RNA polymerase enzyme (Section:9.23). This stock was then distributed amongst the various reaction vessels containing the differing GANP concentrations before the transcription reactions were carried out through addition of T7 RNAP and incubation at 37 °C for 2.5 h.

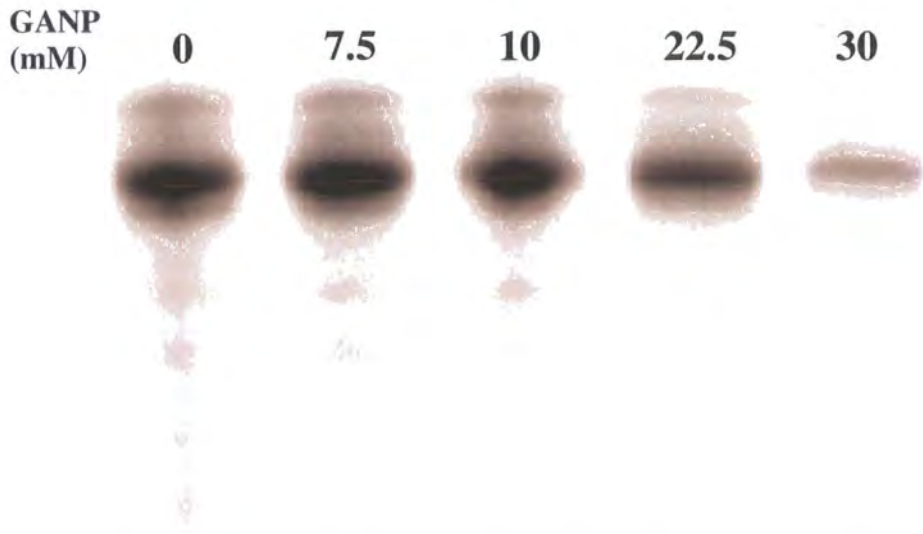


Fig:5.11 IP image of PAGE gel used to isolate ATRib^{TL} RNA transcripts synthesised in the presence of differing GANP concentrations.

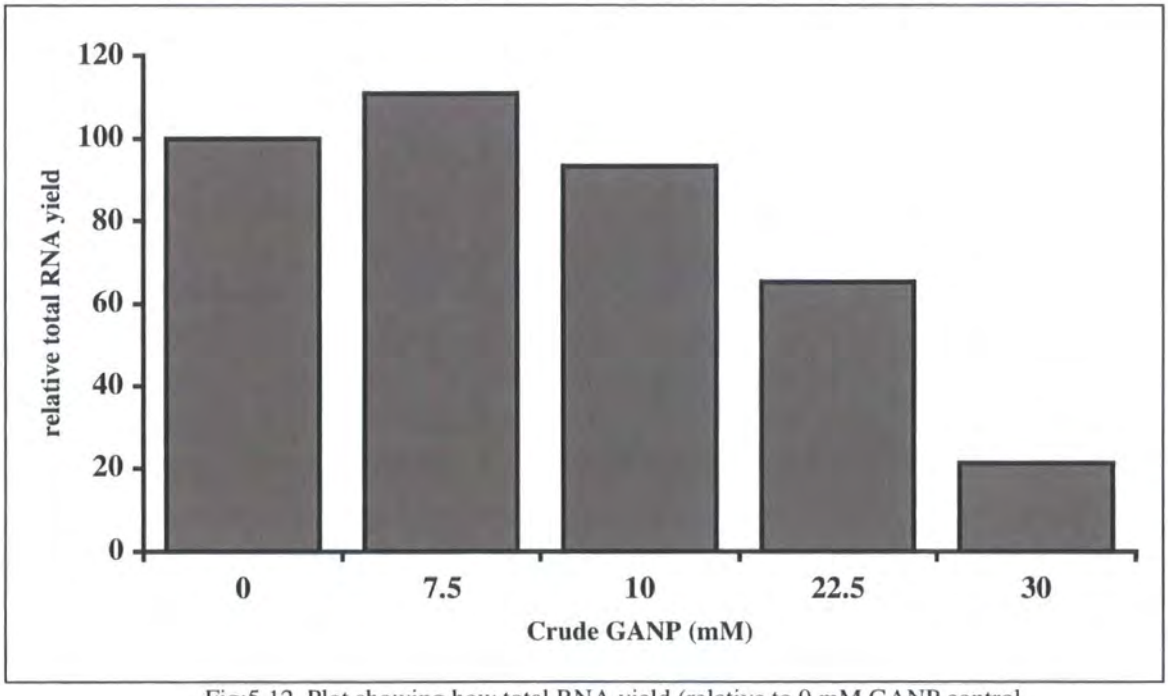


Fig:5.12 Plot showing how total RNA yield (relative to 0 mM GANP control reaction) varies with crude GANP concentration. Data collected from IP image shown in Fig:5.11.

In Fig:5.11–12 a clear trend between GANP concentration and RNA yield can be observed:

- At low concentrations (7.5 mM) RNA yield is increased relative to the 0 mM control
- At higher concentrations (10 – 30 mM) RNA yield is significantly decreased.

Hence, it is apparent that to maximise RNA yield we should use a GANP concentration in the order of 7.5 mM. The question of whether or not a low concentration of GANP would give acceptable levels of incorporation was answered *via* the biotinylation and streptavidin- assay of the isolated ATRib^{TL} transcripts.

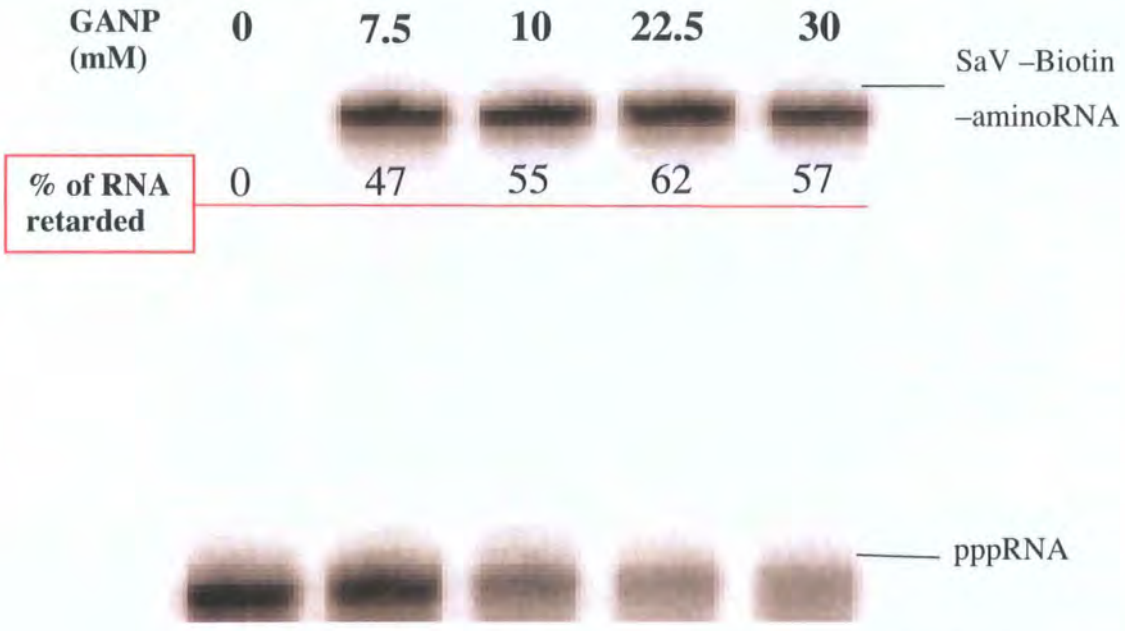


Fig:5.13 IP image of streptavidin-assay. The relative intensities of the slower and faster migrating bands in each lane were used to determine the percentage of GANP incorporation for each ATRib^{TL} sample.

The image (Fig:5.13) above demonstrates how in general an increase in the amount of GANP used in a transcription reaction results in a coinciding rise in retarded RNA in the streptavidin assay. Bands retarded in a streptavidin assay are those aminoRNA transcripts

formed through GANP incorporation that had been subsequently biotinylated and conjugated to streptavidin protein.

5.11 Results of Crude GANP optimization

By collating the results from the transcription and subsequent bioconjugation of the ATRib^{TL} transcripts we can form a meaningful opinion on the optimum reaction conditions for the formation of 5′-AminoRNA using crude GANP.

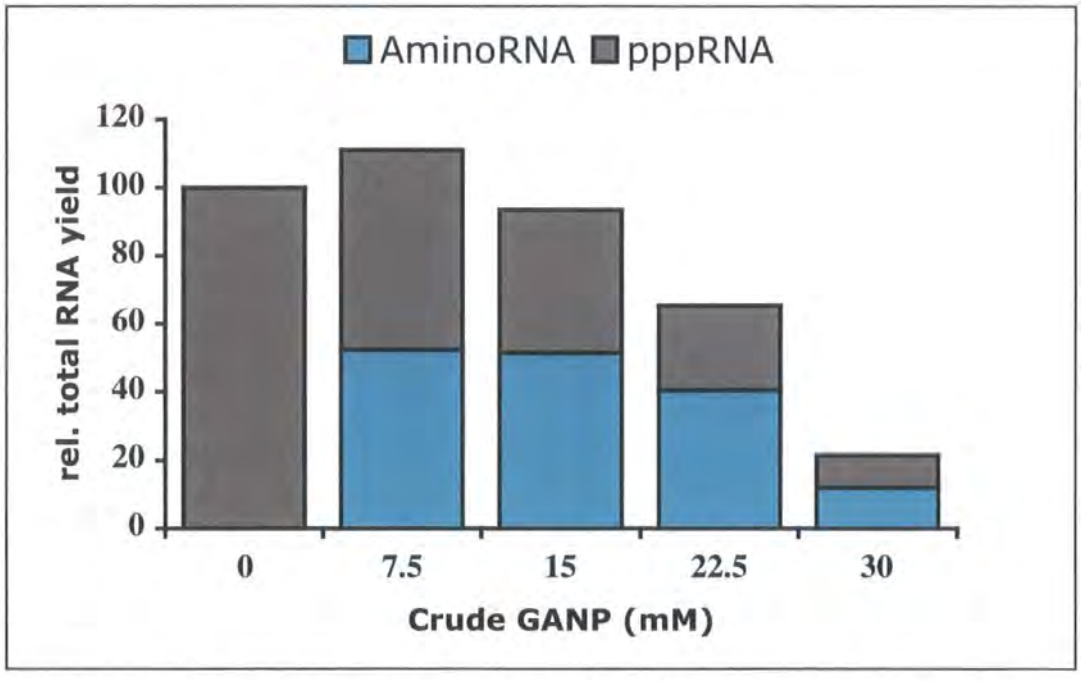


Fig:5.14 Plot of relative total RNA yield vs Crude GANP concentration. The height of each bar corresponds to the relative total RNA yield for each transcription reaction. Each bar is separated into two areas, designating what fraction of the transcripts had been primed with GANP to form AminoRNA.

Crude GANP (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
7.5	111	47
15	93	54
22.5	65	62
30	21	56

Table:5.1 Relative total RNA yields (relative to the 0 mM GANP control experiment) and percentages of GANP incorporation (% of total RNA yield retarded in streptavidin assay) for each concentration of phosphoramidate used.

The plot above (Fig:5.14) clearly demonstrates that variations in total RNA yield and percentage GANP incorporation results in comparable levels of 5'-AminoRNA being produced through the use of 7.5 and 15 mM crude GANP. This result is accounted for as the lowering of total RNA yield with increasing GANP concentration is compensated for by the increase in GANP incorporation level. Transcriptions using higher concentrations of GANP do not display incorporation levels high enough to counter the decrease in total RNA yield.

Through the use of crude GANP sample we were able to make a preliminary assessment of the phosphoramidate's ability to initiate RNA transcription. With a cleaner sample of GANP available to us (see Section 2.58) we were in a position to investigate and compare the behaviour of the purified compound.

5.12 Transcriptions using purified GANP

Crude phosphorylation reaction mixture such as that used in Sections 5.4 onwards was purified by ion-exchange chromatography and precipitation methods as described in Section 2.57, the collected GANP was then used in transcription reactions. Taking a lead from the result gained using crude GANP; we carried out transcription reactions using purified GANP in the range 0–15 mM. We hypothesised that reactions using purified

GANP would result in a lower optimum concentration than that observed for the crude product.

The reasoning for this hypothesis was that the crude phosphorylation reaction mixture contained less than 100 % GANP, therefore the assumed phosphoramidate concentrations for each transcription reaction was higher than the actual concentration present. In addition, the base and salt impurities in the sample could have had a negative effect on the enzyme's functionality. This being said the use of the crude product gave results comparable in efficiency with other novel nucleotide incorporating transcription reactions (see Table:1.35).

Although we attempted to minimise the degradation of our product by working at high pH and low temperatures when possible, the extensive amount of chromatography required to purify the novel nucleotide means some hydrolysis was unavoidable. Hence our purified GANP also contained a small amount of sodium hydroxide that was residual from strong anion exchange chromatography. Unfortunately subsequent gel filtration chromatography and precipitation could not remove 100 % of this impurity and we were forced to accept its presence.

To ensure the impurities present in the sample did not adjust the pH of the transcription reaction above optimum conditions, we adjusted the 100 mM GANP stock solution to pH 8 with 1 M HCl. Encouraged by the results collected when using the crude GANP mixture, which contained much higher levels of salt than the purified sample, we believed the T7 RNAP enzyme would be more tolerant of the small amounts of impurities in the reaction mixture containing purified GANP.

To guarantee the results we observed during this study could be attributed to precise concentrations of GANP, the stock solutions of phosphoramidate were evaluated by UV spectroscopy. This was not possible when using a crude sample of GANP as the residual water and salts present made using ^{1}H NMR spectroscopy to analyse the proportion of Amino–G present extremely difficult. As the amine is also UV active and, as all

modifications were made remote to the guanine chromophore, the amine will possess a very similar if not identical extinction coefficient to GANP, therefore we could not accurately attribute measured UV absorbance to a single species. Through purification of the phosphoramidate we collected detailed ^{1}H NMR spectra, allowing us to determine the exact ratio of GANP: Amino-G present in the sample. This allowed us to use UV spectroscopy to make stock solutions of the required GANP molarity.

5.13 Purified GANP results

A number of transcription reactions were carried out using 0–15 mM GANP, these results were used to determine the average total RNA yields and incorporation levels at each concentration studied. The images below are from a representative transcription reaction and subsequent product analysis *via* bioconjugation. The tables and charts give the average of all comparable results obtained.

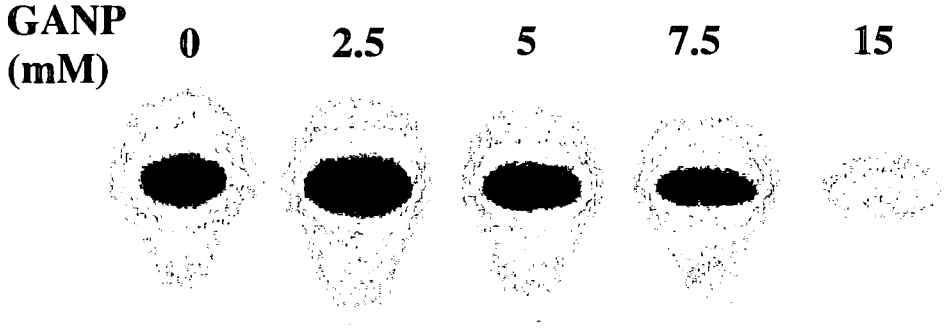


Fig:5.15 IP image of PAGE gel used to isolate ATRib^{TL} RNA transcripts synthesised in the presence of differing GANP concentrations.

The image above (Fig:5.15) displays a trend also evident in our crude GANP investigation, *i.e.* yield is reduced as GANP concentration increases. However, when using purified GANP the reduction in yield is far more exaggerated. It is possible that this trend is due to the more precise GANP concentration in the stock solution, compared to that of the impure sample in which the GANP concentration was given a high estimate.

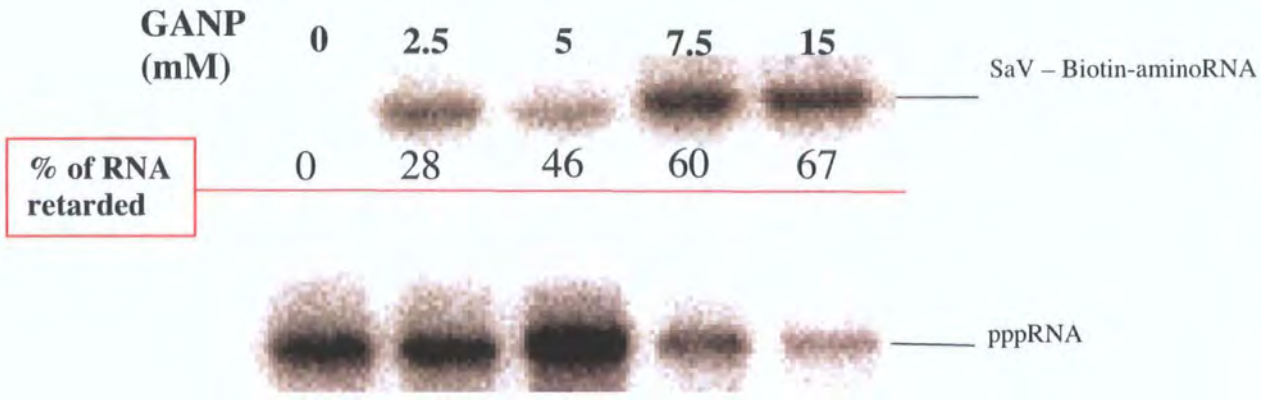


Fig:5.16 IP image of streptavidin-assay of biotinylated RNA transcripts produced in the presence of varying amounts of purified GANP . The relative intensities of the slower and faster migrating bands in each lane were used to determine the percentage of GANP incorporation for each ATRib^{TL} sample.

We observed the expected increase in streptavidin bound, biotinylated RNA as concentration of GANP in the reaction mixtures is increased. To unequivocally confirm that the slower moving band was a result of a streptavidin bound, biotinylated–aminoRNA we ran a gel containing a series of negative controls.

GANP	+	+	+	+	–	–	–	–
Biotin	+	+	–	–	+	+	–	–
SaV	+	–	+	–	+	–	+	–

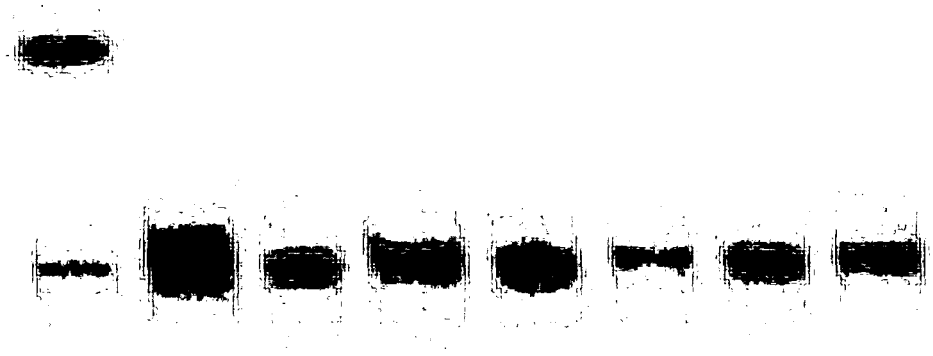


Fig:5.17 Streptavidin-assay control gel. Image confirms that only those RNA transcripts resulting from a (purified) GANP incorporating transcription reaction that have been biotinylated, and subsequently conjugated to streptavidin, produce a slow moving band in the PAGE gel. Other lanes include all other combinations of GANP, biotinylation and streptavidin, none of which yield a slow migrating band.

As the control gel only contains a retarded band in the lane including the biotinylated, GANP incorporating RNA, in the presence of streptavidin. Hence, we can be confident that our assay can be used separate only those transcripts primed with GANP, and therefore allow us (assuming 100% of AminoRNA is converted during the biotinylation reaction) to determine the levels of GANP incorporated in each transcription reaction.

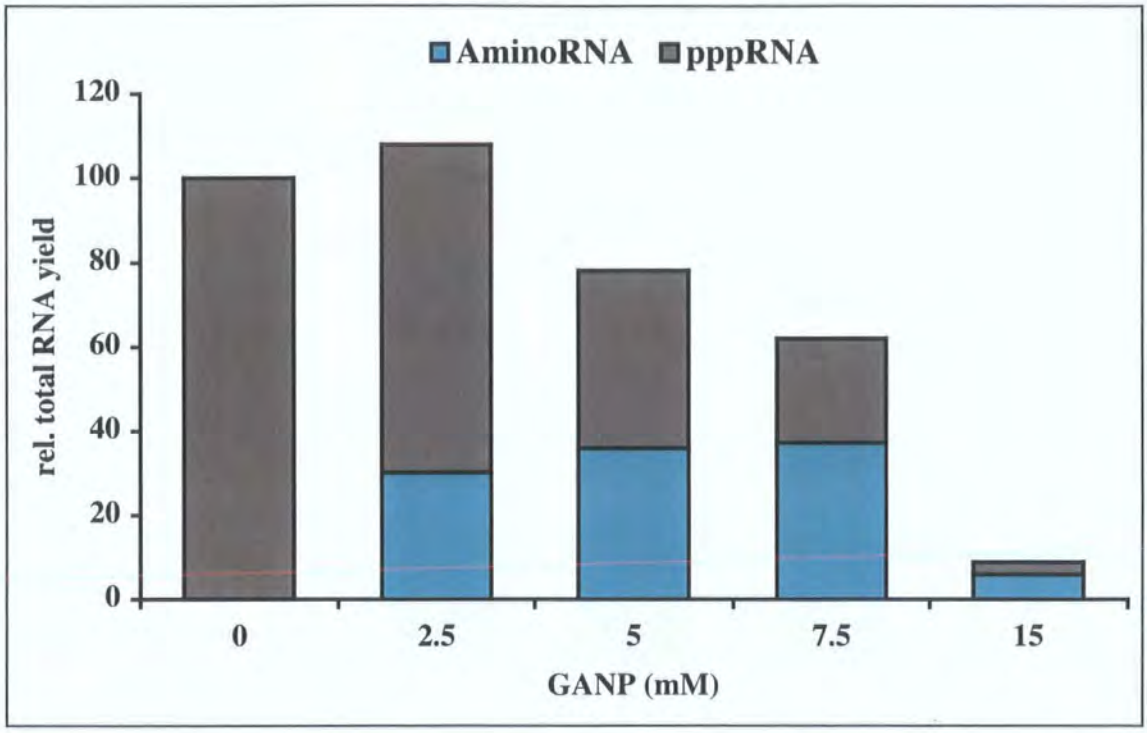


Fig:5.18 Plot of relative total RNA yield vs GANP concentration. The height of each bar corresponds to the relative total RNA yield for each transcription reaction. Each bar is separated into two areas, designating what percentage of the transcripts had been primed with GANP to form AminoRNA.

GANP (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
2.5	108	28
5	78	46
7.5	62	60
15	9	67

Table:5.2 Relative total RNA yields and percentages of GANP incorporation for each phosphoramidate concentration used.

Through comparison of the total RNA yield and level of incorporation observed when using crude or purified GANP, we can identify a number of obvious differences.

- Levels of incorporation at each concentration are higher when using a purer sample.

- The apparent inhibition of RNA synthesis is much more pronounced when using a purer GANP sample.

The increase in levels of incorporation when using the purer GANP fits our hypothesis, as we know these transcriptions contained more novel nucleotide than the corresponding reactions using a crude sample. What is surprising is the dramatic fall off in RNA production when using a relatively small amount of GANP. Comparing the 15 mM result from the purified and crude samples:

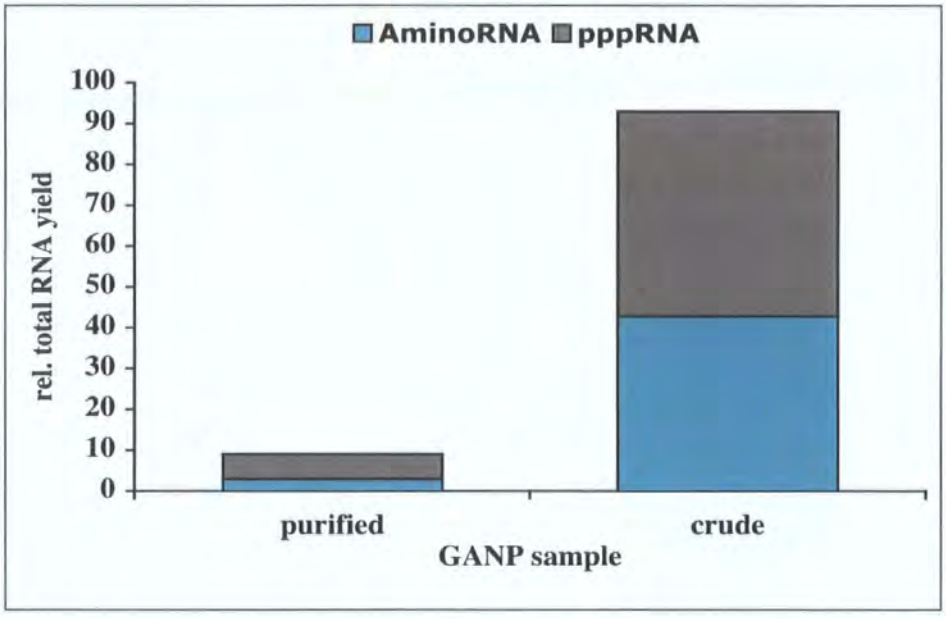


Fig:5.19 Plot of relative total RNA yield vs GANP source (15 mM), illustrating the surprisingly different results observed.

Although there is a slight increase in the level of GANP incorporation when using a purer sample (from 54 to 67 %) the reduction in yield is not proportional.

5.14 Study using reduced GTP concentration

As with all novel nucleotide initiators the level of incorporation of GANP into RNA transcripts is based on competition with GTP to bind to the enzyme active site and the DNA template. It has been shown that it can be beneficial to sway this competition in favour of the synthetic initiator through not only increasing the concentration of novel

nucleotide used, but also by reducing the amount of GTP present in the reaction. For example, ATP, UTP and CTP could all be used in 1.25 mM concentration, while GTP is used at 0.31 mM (1/4 equivalents) and GANP is used at 2.5 – 15 mM. With a higher ratio of GANP: GTP we would expect to observe increased levels of GANP incorporation relative to a reaction with a full quotient of GTP.

However, as illustrated earlier in this chapter the introduction of excess initiator to the transcription reaction can result in a reduction of total RNA yield through competition for template binding during the chain elongation phase of transcription. With a higher ratio of GANP: GTP than previously used, this competition is also swayed in the novel nucleotides favour, resulting in a decreased total RNA yield. A compromise must be met between the two factors of RNA yield and GANP incorporation. In an attempt to improve incorporation efficiency beyond that observed in the previous investigation we carried out transcription reactions with a reduced concentration of GTP (0.31 mM) and various GANP concentrations. The results from this experiment are shown below:

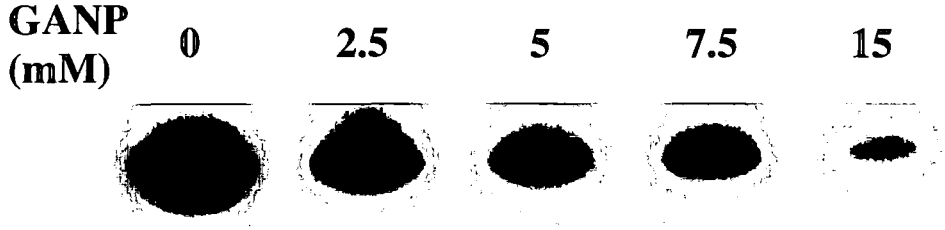


Fig:5.20 IP image of PAGE gel used to isolate ATRib^{TL} RNA transcripts synthesised in the presence of differing GANP concentrations and a reduced GTP concentration.

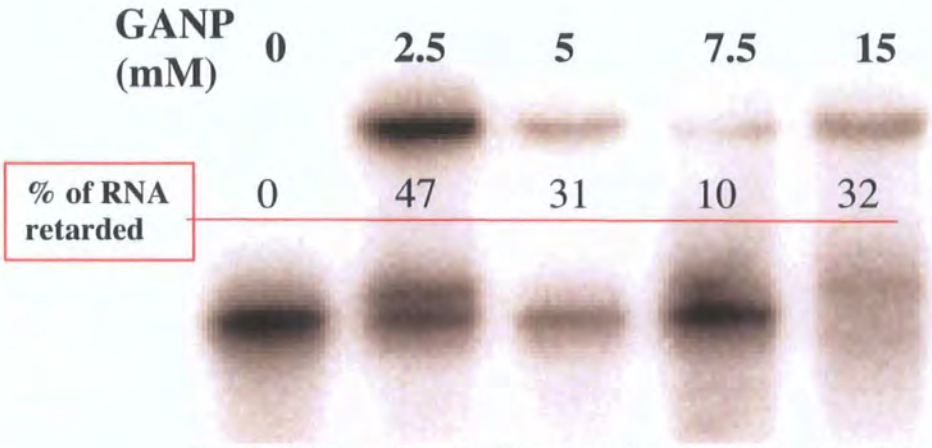


Fig:5.21 IP image of streptavidin-assay. The relative intensities of the slower and faster migrating bands in each lane were used to determine the percentage of GANP incorporation for each ATRib^{TL} sample.

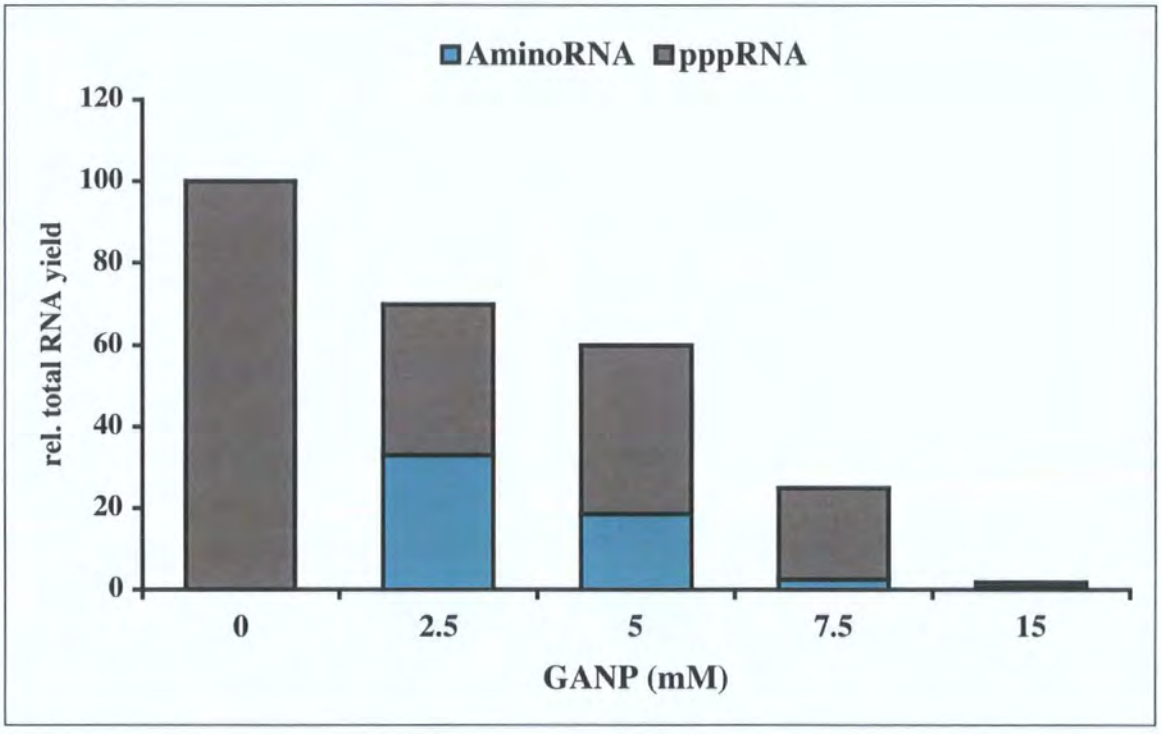


Fig:5.22 Plot of relative total RNA yield vs GANP concentration. The height of each bar corresponds to the relative total RNA yield for each transcription reaction. Each bar is separated into two areas, designating what percentage of the transcripts had been primed with GANP to form AminoRNA.

GANP (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
2.5	70	47
5	60	31
7.5	25	10
15	2	32

Table:5.3 Relative total RNA yields and percentages of GANP incorporation for each phosphoramidate concentration used.

Our prediction that the RNA transcribed in the presence of a reduced GTP concentration would have display an increased level of GANP incorporation with a coincident decrease in total RNA yield, was not borne out by experimental results. Hence, using a decreased amount of GTP in the reaction mixture does not provide a more efficient method of synthesising aminoRNA compared to using a standard ratio of NTPs *i.e.* 1:1:1:1.

5.15 Conclusions of GANP study

Having completed our research into the use of GANP in transcription reactions it is pertinent to review our initial objectives and assess to what degree we have achieved them.

- Establish whether GANP can be used to prime T7 RNA polymerase mediated transcription reactions, resulting in the synthesis of 5′-AminoRNA.
- If successful, can this reaction be optimised to maximise 5′–AminoRNA production.

Our primary target has certainly been met as it is evident GANP is as active an initiator as we had hoped. However, the optimum conditions for the use of GANP are more complex. Comparison of total RNA yields and levels of incorporation resulting from the use of crude and purified samples of GANP give quite different results. Intuitively we expected poorer results when using the impure sample, however this was not the case.

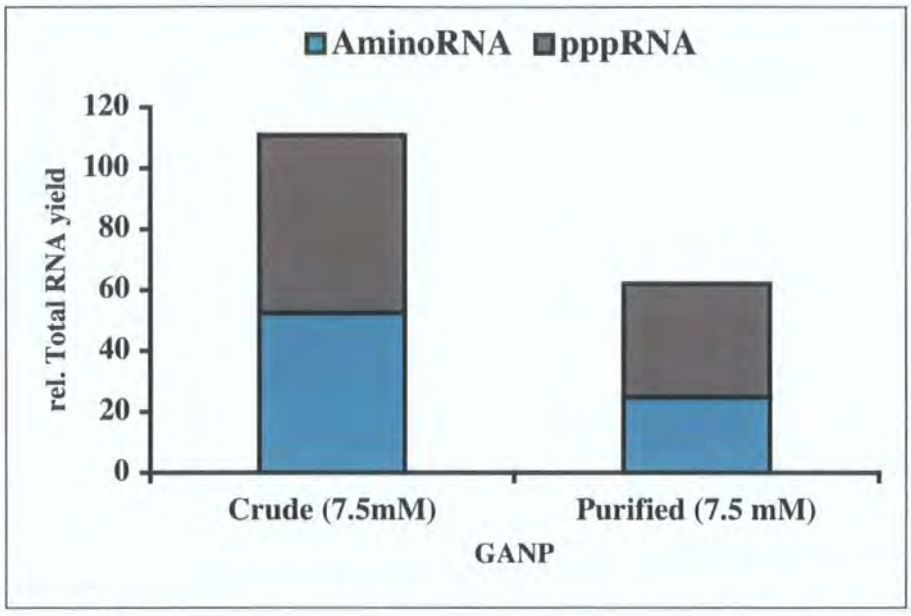


Fig:5.23 Comparison of optimum results gained using crude and purified GANP.

The use of crude GANP:

- Gave high total RNA yields,
- Required high concentrations to inhibit RNA production.
- Resulted in incorporation levels on a par with existing GSMP, GMPS technology and above those of literature methods using Amino–G.

Purification of the crude sample *via* LC and selective precipitation methods gave a much cleaner sample from which we were able to collect characteristic ^{1}H, ^{13}C, ^{31}P NMR spectra. With fewer impurities in the reaction mixture (some Amino-G, phosphate and hydroxide was still present in the sample) we expected to see both a good total RNA yield and high levels of incorporation. Instead what was observed experimentally was:

- Inhibition of RNA production even at low concentrations.
- Marginally higher levels of incorporation than when using the crude sample.

The higher level of transcription inhibition observed when using purified GANP is likely owing to increased competition between GANP and GTP in the elongation stage of polymerisation. As the crude GANP contains a lower amount of phosphoramidate there

is less competition and therefore less inhibition is observed. The approximately equal levels of incorporation (55–67 %) observed when using either crude or purified GANP seems to indicate that a limit has been reached on the phosphoramidates ability to compete with GTP to initiate transcription. An increase in GANP concentration once this limit has been reached does not raise the level of incorporation, but does reduce the total RNA yield.

Taking these results into consideration we can theorise uses for both the crude and purified GANP. If a large quantity of 5′-AminoRNA is required, irrespective of the amount of triphosphate terminated RNA also transcribed, then use of the crude GANP sample would be advised. Not only would the user avoid the extensive steps necessary to purify the GANP before use, but would benefit from a greater total RNA yield and good level of incorporation.

When the maximum level of GANP incorporation must be achieved, irrespective of total RNA yield *e.g.* in the transcription of a population of potential ribozymes containing 10^{14} randomly sequenced transcripts, any of which could be active, the use of the purer GANP could be justified. Though whether the marginally higher level of transcription is worth the time spent purifying the GANP would be at the discretion of the individual researcher.

The next chapter will discuss a third option for the enzymatic synthesis of 5′-AminoRNA, the use of Amino–G in transcription reactions. This vein of research opened up to us owing to observation of the pH dependent solubility of Amino–G during the synthesis of GANP, and resulted in the need for a phosphoramidate masking group being brought into doubt.

6.0 Amino–G Transcription Study

6.1 Targets for the use of 5′-amino-5′-deoxyguanosine

This chapter will detail our studies into the ability of Amino–G to initiate T7 RNA polymerase transcription reactions, and in doing so synthesising 5′–AminoRNA. As previously discussed (Section 1.33) the use of Amino–G in transcription reactions by other research groups had resulted in a low level of incorporation (20%) thought to be owing to the low solubility of Amino–G in the transcription reaction. Our original intention was to only use Amino–G as a standard for our investigation into the efficiency of GANP in transcription reactions. However during the synthesis of GANP it became apparent that the amine was more soluble in water than had been previously observed, therefore we extended our research to include the use of Amino–G as an initiator for transcription reactions. We believed an increase in the solubility of Amino–G would have a pronounced effect on its level of incorporation in comparison to previously reported results. After investigating the limits of the amine's solubility by UV spectroscopy we carried out a series of transcription reactions using Amino–G as an initiating nucleoside. Through analysing the products of these reactions we determined the total RNA yield and level of amine incorporation achieved when using a range of Amino–G concentrations. The collected data was used to prescribe optimum conditions to maximise the production of 5′–AminoRNA when using Amino–G. Transcription reactions using Amino–G and a reduced GTP concentration were undertaken, resulting in a lowered total RNA yield and an excellent level of amine incorporation (88 %).

6.2 Analysis of Amino–G solubility

As discussed in Chapter 2, we believe the increase in the amine's solubility was due to a deprotonation brought about by the presence of base in the aqueous solution. The consequential increase in the polarity of the novel nucleoside would lead to a heightened solubility in polar solvents.

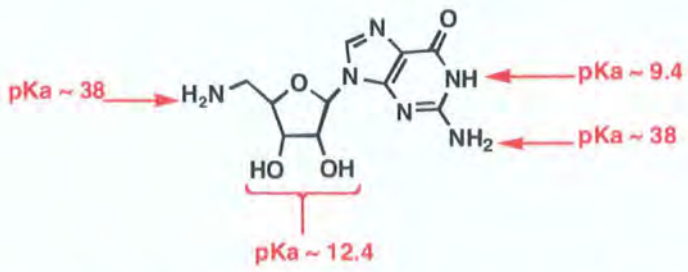


Fig:6.1 pK_a values of the ionisable groups present in Amino–G[32].

The proton most likely to be abstracted is that of the amine at position 5 on the guanine base ($pK_a \sim 9.4$). For this increase in performance to be possible the solubility of the amine should be increased to a level close to that of other novel nucleotide initiators (*e.g.* GMPS, GSMP[26,28,31]). Therefore before carrying out a study into the ability of Amino–G to initiate transcription we first used UV spectroscopy to assess the solubility of the amine at high and moderate pH. To ascertain whether the limits of Amino–G solubility were broad enough to use the amine in transcription reactions we designed a brief UV study to analyse the concentration of Amino–G in a number of solutions at various pH levels. If our hypothesis were correct we would observe a reduced concentration of Amino–G present in the lower pH solutions. If we could not make solutions of high enough concentration we may not have been able to access the high levels of incorporation observed for GANP and other novel nucleotides. As we were interested in linking the solubility of Amino–G to the low incorporation levels observed by Suga *et al*, one of the solutions we wanted to study was T7 reaction buffer adjusted to pH 9, a mimic of their transcription reaction conditions[45].

6.3 Methodology of the UV spectroscopy study

The method we followed is detailed below:

Enough Amino–G was added into two sets of 5 tubes to make 100 μL of the following concentrations 12.5, 25, 37.5, 75, 112.5 mM. We chose these values, as they are the concentrations of the undiluted stocks used in our transcription reaction experiments, and therefore represent the range of concentrations we would require when studying the transcription performance of Amino–G. Each set of tubes would have either a solution of

T7 buffer, adjusted to pH 9, or sodium hydroxide (100mM) added to it. All the tubes were then agitated for an hour to allow time for the maximum amount of amine to dissolve. The concentration of Amino–G in each solution would then be determined by UV spectroscopy and the results compared. As a final step the pH of the NaOH solutions would be reduced to pH 9 and the concentration of Amino–G reassessed.

Set A

To one set of tubes (A) we added 100 μL of T7 buffer at pH 9. A photograph was taken of the tubes (Fig:6.2 A.1) before they were mechanically shaken for 1 h at ambient temperature.

The solutions were then centrifuged and the liquid removed. The concentration of Amino–G in each solution was then assessed by UV spectroscopy.

Set B

To one set of tubes (B) we added water (90 μL). A photograph was taken of the tubes (Fig:6.3 B.1) before addition of NaOH (10 μL, 1M) and mechanical shaking for 1 h at ambient temperature.

After shaking another photograph was taken showing the full dissolution of all solids.

The solutions were then centrifuged and the liquid removed. The concentration of Amino–G in each solution was then assessed by UV spectroscopy.

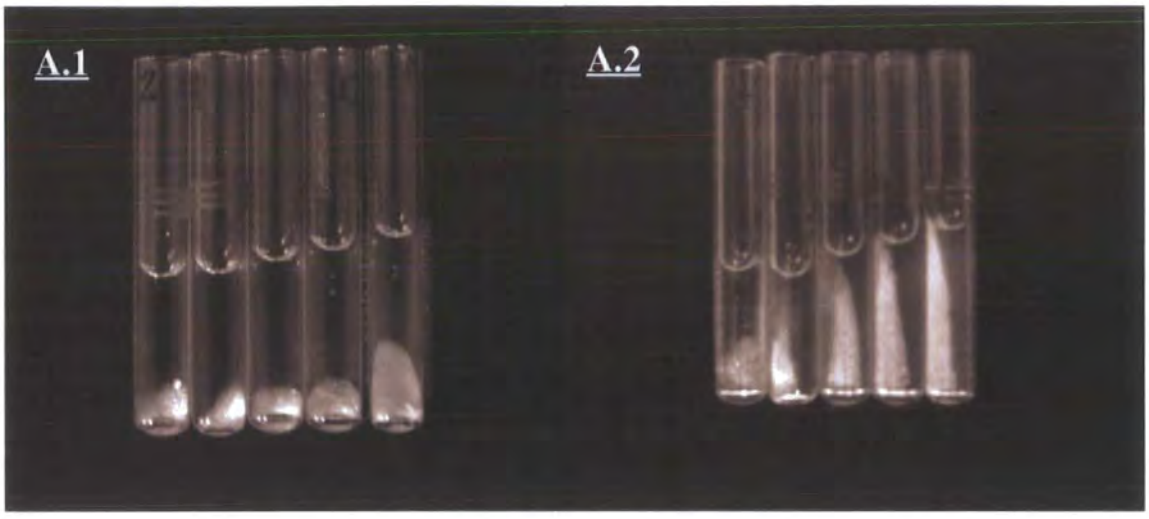


Fig:6.2 **A.1:** From left to right 12.5, 25, 37.5, 75, 112.5 mM Amino–G, in 100 μL pH 9 T7 buffer at time = 0. **A.2:** Samples as A.1, after 1 h mechanical shaking, no visible dissolution of solid occurred.

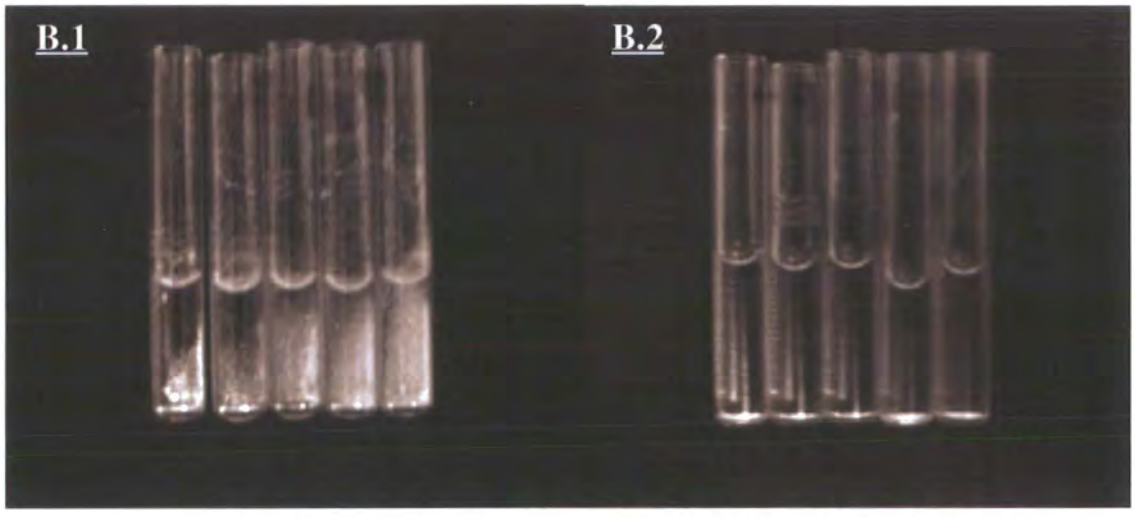


Fig:6.3 **B.1:** From left to right 12.5, 25, 37.5, 75, 112.5 mM Amino–G, in 90 μL DEPC water at time = 0. **B.2:** Samples as A.1, after addition of NaOH (10 μL,1M) and mechanical shaking for 1 h. Complete dissolution of solid was observed.

In the case of set B (NaOH solution) complete dissolution of solid took place directly upon addition of the NaOH. To each tube we added a quantity of HCl to reduce the pH of the solution to 9. Upon addition of the acid some precipitate was observed (Fig:6.4).

Fig:6.4 **B.3:** Samples as per Fig:6.14. The reduction of the pH level to 9 with 1 M HCl caused precipitation of a white solid. More solid was visible in the more concentrated amino-G solutions

If the concentration of Amino–G was dependent on the pH of the solution, and not affected by the route taken to achieve that pH level, we would expect the concentration of the pH 9 solutions to be comparable to those determined for the T7 buffered solution. If the route to the pH level were important however, we would expect a difference in the observed concentrations.

From a thermodynamic viewpoint the re–protonation of the solubilised Amino–G renders it less soluble in the pH 9 solution. As a significant proportion of the amine remains in solution it would be reasonable to assume that not all of the Amino–G re–protonates when the pH is reduced from 13 to 9. As the species to be re–protonated has a pK_a of 10, indicating that at pH 9 the majority of amine would be protonated and insoluble, we can hypothesise that a kinetic barrier to the protonation exists, the nature and magnitude of which cannot be ascertained from the data collected so far. Without further experimentation we can only hypothesise that such a kinetic barrier exists and that it is this factor which inhibits the precipitation of all the Amino–G from the pH 9 solution.

Section 6.14 discusses further experiments that could be carried out to shed light on this and other questions arising from the pH dependent solubility of Amino–G.

6.4 Results of solubility study

As no extinction coefficient is available for Amino–G we used the ε_{260} of guanosine, 12080 M^{-1}. We felt this value was still applicable to the amine, as the alterations to the guanosine compound were made at a position remote to the UV active guanine group. A slight variation in the ε_{260} would lead to a negligible change to the concentration values acquired.

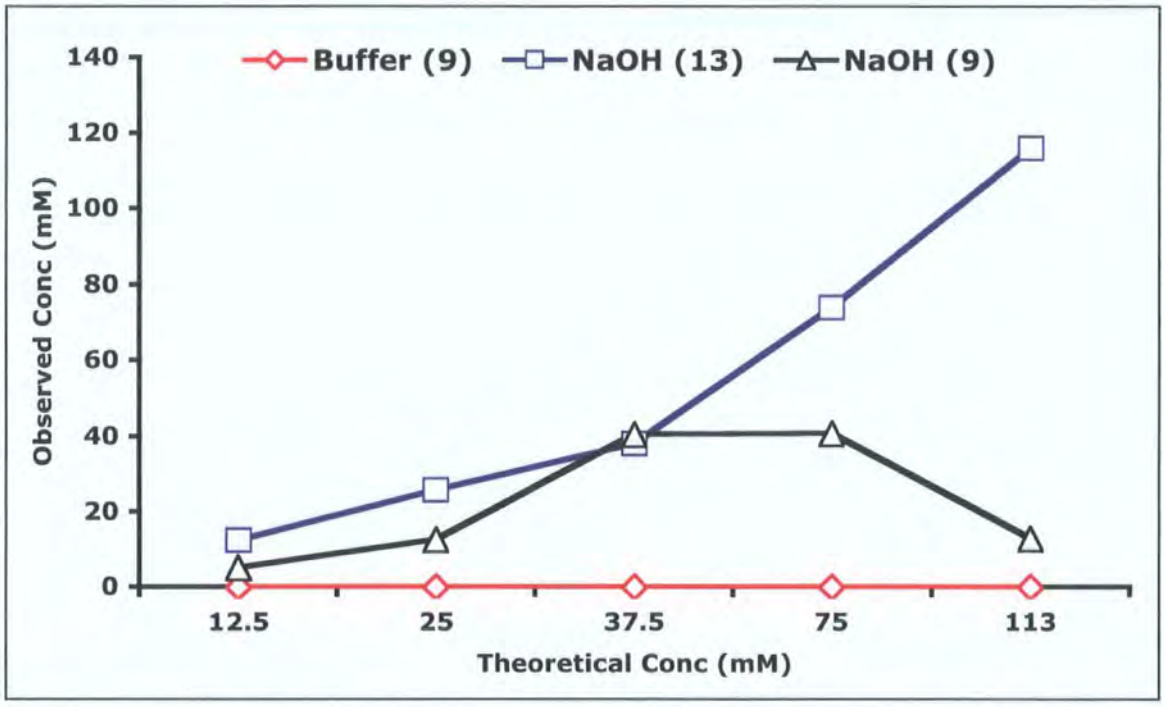


Fig:6.5 Plot showing theoretical Amino–G concentration vs observed concentration as determined *via* UV spectroscopy.

Amino–G concentration (mM)			
Theoretical	**Buffer (9)**	**NaOH (13)**	**NaOH (9)**
12.5	0	12.4	5.2
25	0	25.5	12.5
37.5	0	38	40.3
75	0	73.7	40.6
112.5	0	116.1	12.7

Table:6.1 Data showing theoretical vs observed concentrations of Amino–G in T7 buffered, and high pH solutions.

We observed a very significant difference in Amino–G concentrations between the buffered and aqueous solutions. Unexpectedly no Amino–G was found to have dissolved in the T7 buffer, from the published levels of Amino–G incorporation we had expected to observe a low concentration of the amine in these solutions. This inconsistency was attributed to the differing temperatures that our solubility study and the transcription reaction conditions used by Suga *et al* employed[45]. Our experiment was carried out at room temperature (~20 °C), however, transcription reactions are incubated at 37 °C as standard, conditions which would aid the dissolution of solid. Although unhelpful in discerning the true concentration of the amine present in a pH 9 solution under transcription reaction conditions, our results can still be used to illustrate the pH dependent solubility of Amino–G.

In accordance with our expectations the highest concentrations were observed in the pH 13 solutions, all of which were found to be near consistent with the theoretical concentration values. Deviations from the desired concentrations are likely to be owing to errors in weighing the small quantities of Amino–G required for the study (~2-5 mg per sample). As predicted we observed a reduction in Amino–G concentration as the pH of the aqueous solution was lowered. This result demonstrates that not only could we make Amino–G solutions of suitable concentrations for use in a transcription study, but also corroborates the hypothesis that the source of Amino–G's poor performance in past studies was owing to its low solubility.

6.5 Aims and objectives

The aims of our study into the ability of Amino–G to initiate transcription reactions were very similar to those in our research into the use of GANP:

- To establish whether the use of Amino–G solubilised in a weak $NaOH_{(aq)}$ solution could prime T7 RNA polymerase mediated transcription reactions with more success than previous methodologies have reported.

- To optimise the amount of Amino–G in the transcription reaction to maximise 5′–AminoRNA production, taking into account both the total RNA yield and level of amine incorporation.

At the onset we realised that when using relatively high concentrations of Amino–G the sodium hydroxide in the stock solution may overwhelm the transcription reaction's buffer capability. Therefore an additional aspect of the transcription reactions required our attention:

- To determine what effect, if any, the hydroxide in the amine stock has upon the total RNA yield and the level of Amino–G incorporation.

It is known that enzymes function best at a distinct pH level determined by the structure and function of the enzyme, we would therefore expect that deviations from the optimum pH through addition of excess base (or acid) would result in a reduction in performance. Past studies into the pH tolerance of T7 RNA polymerase have found a 20 % reduction in total RNA yield when the pH of the reaction mixture was altered +/– 1 units from the optimum of 8.

6.6 Existing methodologies for Amino–G use

Attempts have been made to use 5′-amino-5′-deoxyguanosine (Amino–G) to prime RNA sequences, providing a more direct route to 5′–AminoRNA than through the use of GANP. However, published research utilizing T7 RNA polymerase mediated transcription reactions quoted a disappointing maximum incorporation level of 20%,

inefficient when compared to other novel nucleosides that average an incorporation level of ~ 70 %. The low levels of incorporation were believed to be owing to the poor solubility of Amino–G in transcription reaction mixtures, thereby reducing the amines availability for initiating transcription. As our solubility study above illustrates (Section 6.2) the solubility of Amino–G is pH dependent, therefore through the use of a basic solution to pre–dissolve the amine we hope to increase its availability in the reaction mixture.

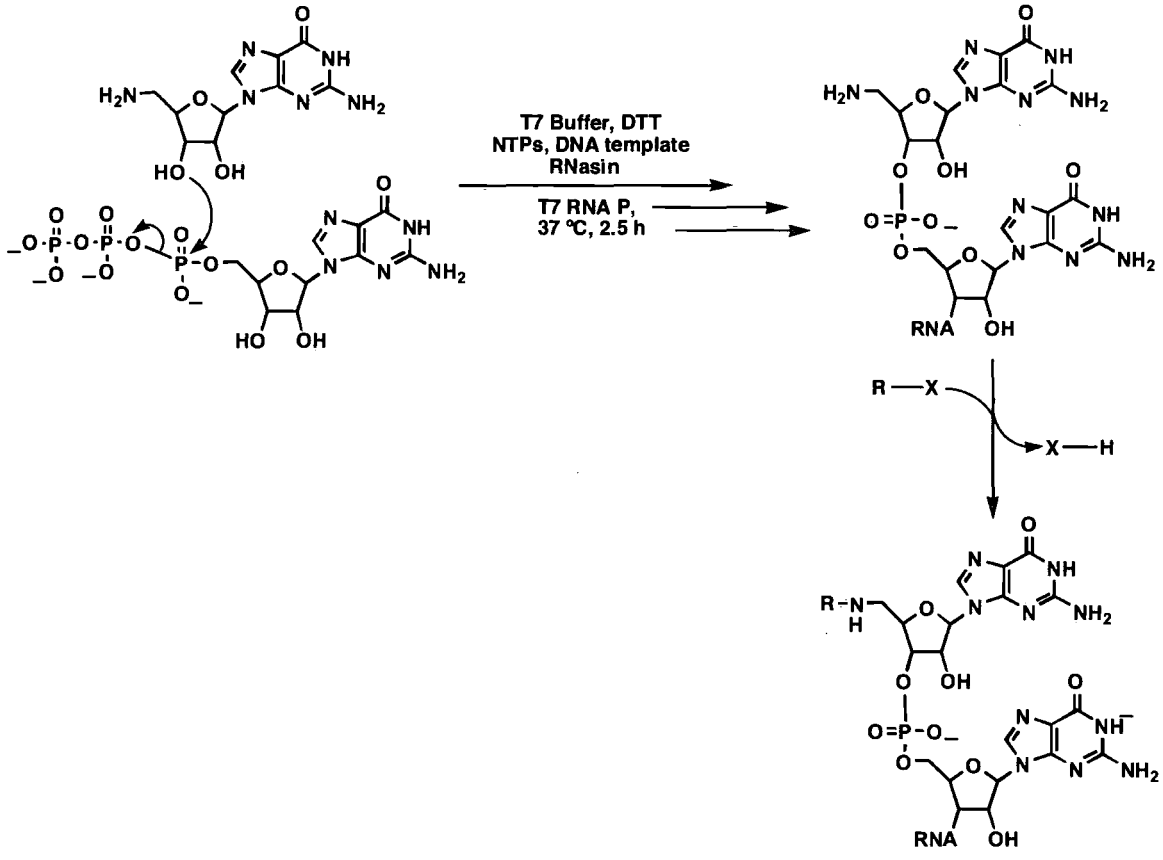


Fig:6.6 Suggested strategy for use of Amino–G in T7 RNAP mediated transcription reactions, resulting in the synthesis of 5′-AminoRNA.

6.7 Factors affecting the use of Amino–G

Competition with GTP for binding during initiation and elongation phases was as significant to the performance of Amino–G as it was for GANP, hence we devised a similar strategy to determine the optimum concentration of Amino–G required for

maximum production of 5'-AminoRNA. That is: the analysis of transcription reaction products resulting from the use of differing Amino–G concentrations.

Through increasing the concentration of Amino–G in transcription reactions we will also unavoidably increase the pH of the reaction mixture, owing to the hydroxide present in the stock solution. The buffered conditions under which pH sensitive enzymes such as T7 RNA polymerase are used should limit the effect of the added base. We expected that the more the pH deviated from the optimum value of 8 the lower the observed activity of the enzyme. Owing to the low concentrations of Amino–G and base being added, we hypothesised that an optimum level of incorporation could be determined at a point where total RNA synthesis was still high yielding.

6.8 Amino–G in transcription reactions

Transcription reactions were carried out *via* the method previously described in the study of GANP (see Chapters 3–4), the products of which were purified through the use of:

1. Zeba Desalt spin columns, to remove excess Amino–G from the reaction mixture,
2. PAGE to separate full length ATRib^{TL} transcripts from aborted transcripts and other reagents/ by–products
3. Excision and elution of transcripts from the acrylamide gel and subsequent precipitation and dissolution of isolated RNA chains.

Upon isolation of the ATRib^{TL} transcripts, biotinylation of the aminoRNA chains present was achieved using the same amine-specific reagent and methodology as used with GANP primed RNA (sulfo-NHS-biotin).

6.9 Results of Amino–G incorporating transcription reactions

Results collected when using low concentrations of Amino–G in transcription reactions revealed an increase in total RNA yield in excess of any similar effect observed in our GANP study. Correspondingly less inhibition of RNA production was observed across the concentration range at which GANP had been studied. The increased total RNA yield

allowed us to investigate the use of higher concentrations of Amino–G than was possible in the case of the relatively inhibitory GANP. The IP images below are from a representative set of transcription reactions using a range of Amino–G concentrations.

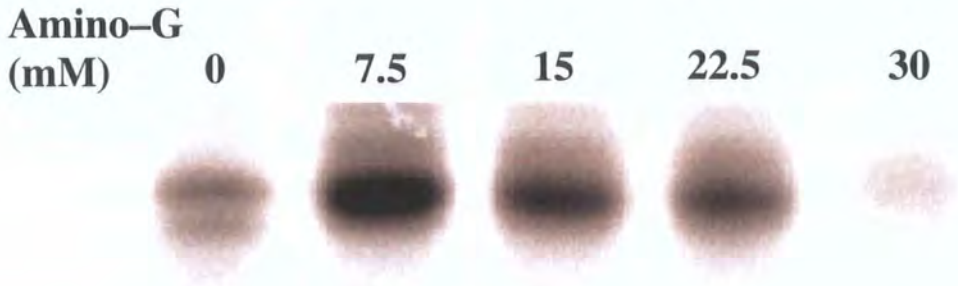


Fig:6.7 IP image of PAGE purification plate. Total yield of ATRib^{TL} transcripts can be seen to vary across the concentration range of Amino–G used. Significantly Amino–G can be seen to stimulate transcription, resulting in higher total RNA yields at 7.5, 15 mM reactions than at 0 mM.

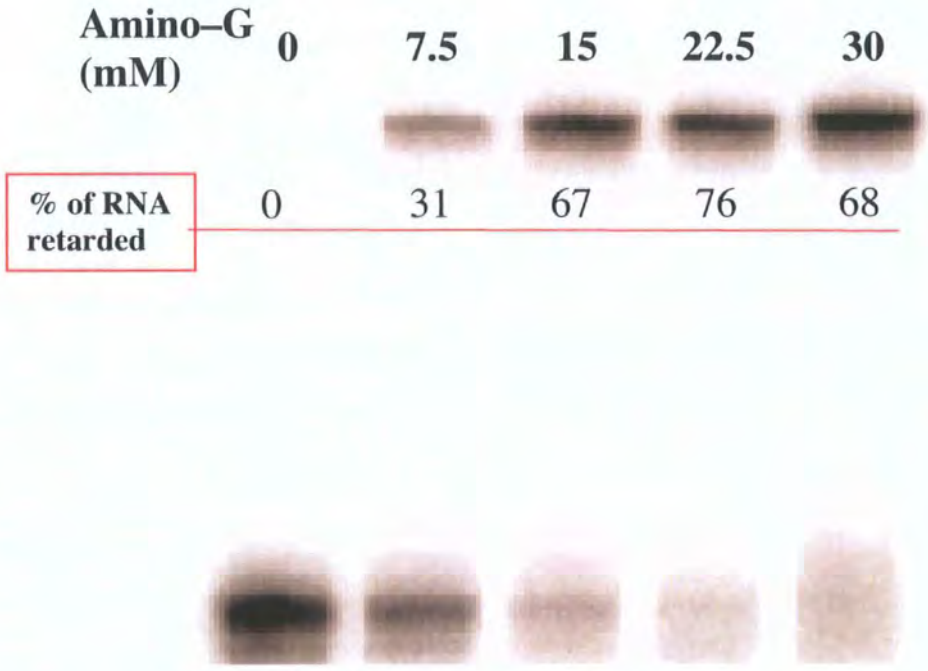


Fig:6.8 IP image of PAGE plate showing results of AminoRNA streptavidin assay . The image shows separation of two radio labelled RNA transcript populations. The transcripts only differ in that one has been primed with Amino–G, thus incorporating an amine at the 5′-terminus, while the other was not.

The degree of biotinylation observed was higher than that observed in literature examples or those reactions undertaken containing either purified or crude GANP samples (Fig:6.8). This result fits our hypothesis that enhancing the solubility of Amino–G would have a significant effect on its level of incorporation. We felt it prudent to run a control gel to confirm the slower migrating bands in the streptavidin assay were indeed owing to an interaction between streptavidin and the biotinylated 5′–AminoRNA, rather than interactions with any other labelled species.

Amino –G	+	+	+	+	–	–	–	–
Biotin	+	+	–	–	+	+	–	–
SaV	+	–	+	–	+	–	+	–

Fig:6.9 Streptavidin-assay control gel. Image confirms that only those RNA transcripts resulting from an Amino–G incorporating transcription reaction that have been biotinylated, and subsequently conjugated to streptavidin, produce a slow migrating band in the PAGE plate.

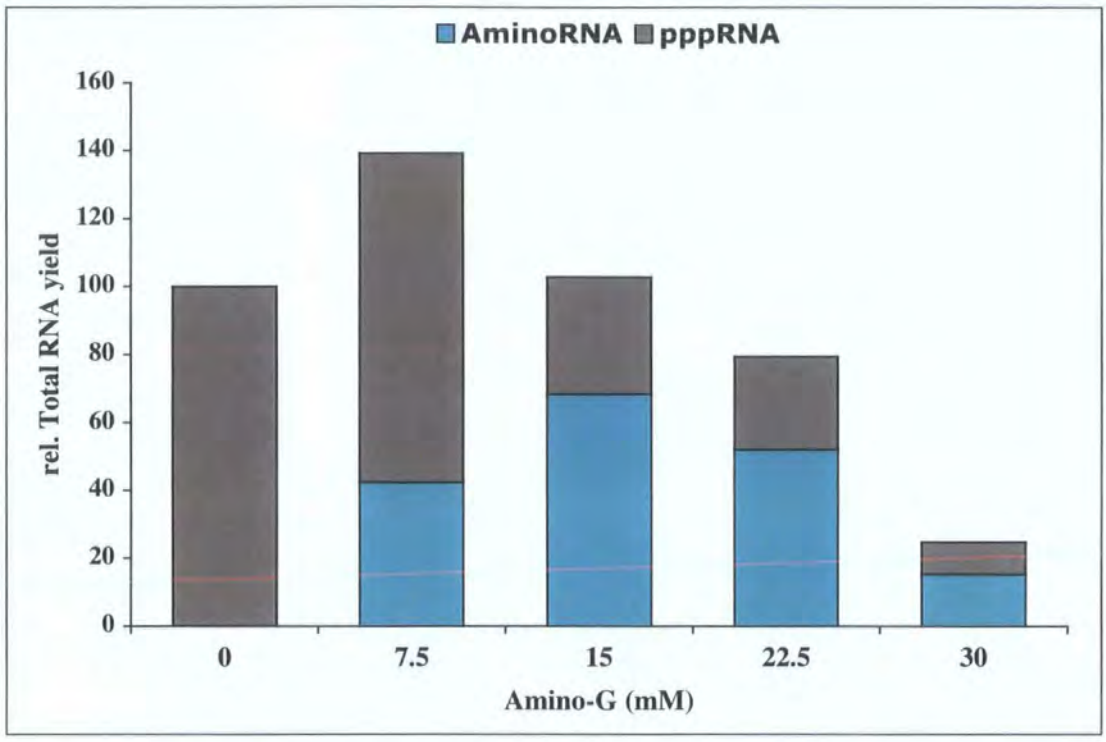


Fig:6.10 Plot of relative total RNA yield *vs* Amino–G concentration. The height of each bar corresponds to the relative total RNA yield for each transcription reaction. Each bar is separated into two areas, designating what percentage of the transcripts had been primed with Amino–G to form 5′–AminoRNA.

Amino–G (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
7.5	139	31
15	102	67
22.5	80	76
30	25	68

Table:6.1 Relative total RNA yields and percentages of amino-G incorporation for each amine concentration used.

The use of Amino–G at 7.5 and 15 mM caused an increase in total RNA yields in comparison to the control reaction. The use of higher concentrations resulted in a reduced total RNA yield, however the inhibition observed was not as drastic as that observed for the corresponding concentrations of GANP.

The comparison to GANP was not entirely favourable as Amino–G appears to initiate RNA transcription less efficiently, this drawback appears to be negated through our ability to use Amino–G in higher concentrations, giving an overall higher amount of 5′-AminoRNA for a given reaction volume. We determined the optimum concentration of Amino–G to use in transcriptions was 15 mM, yielding approximately the same amount of RNA as the 0 mM control reaction, with 67 % of transcripts initiated by the amine to produce 5′–AminoRNA.

6.10 Transcription reactions using a reduced GTP concentration

As in our research into GANP we felt it was prudent to investigate the effect of a reduced GTP concentration on the products of transcription reactions, thereby enhancing the incorporation levels of Amino–G (Section 5.14). Reagents used and the protocol followed for the reactions was identical to those 'standard' transcriptions used elsewhere in this chapter, the only variable being the concentration of GTP used. The images and tables below give the results of this experiment.

Fig:6.11 IP image of PAGE purification plate. Total yield of ATRib^{TL} transcripts can be seen to vary across the concentration range of Amino–G used. * control reaction using 1.25 mM GTP, all other reactions used 0.3 mM (1/4 the standard amount).

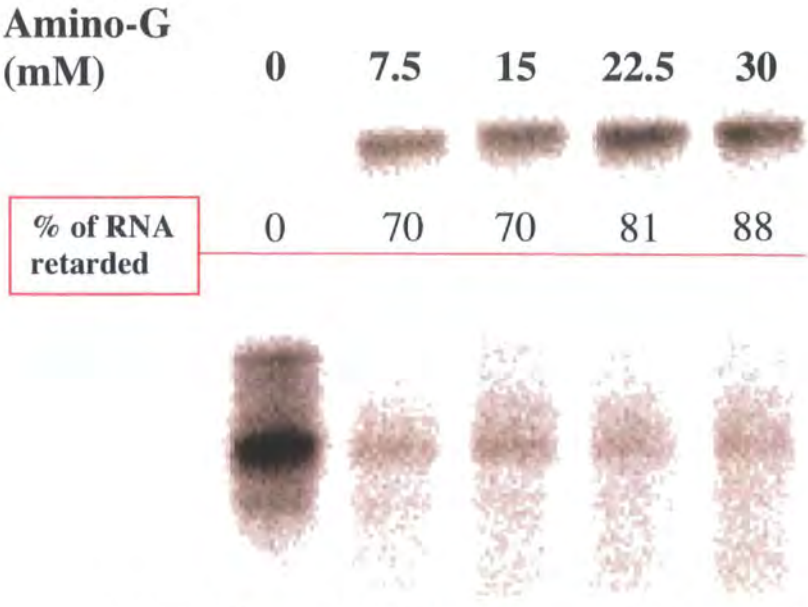


Fig:6.12 IP image of PAGE plate showing results of a streptavidin assay. The image shows separation of two radio labelled RNA transcript populations. Comparison of the two bands intensities allowed the level of Amino–G incorporated into the RNA population to be qualified.

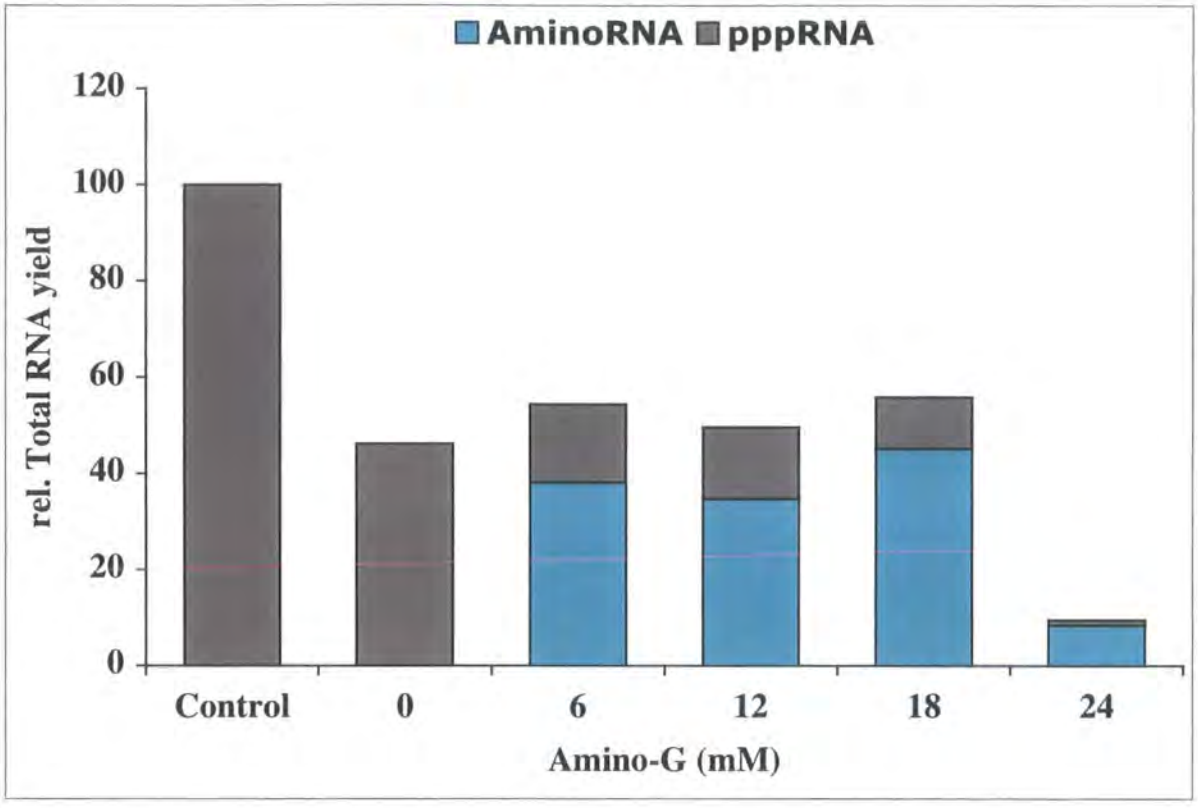


Fig:6.13 Plot of relative total RNA yield vs Amino–G concentration. The height of each bar corresponds to the reative total RNA yield for each transcription reaction. Each bar is separated into two areas, designating what percentage of the transcripts had been primed with Amino–G to form 5′–AminoRNA. The control reaction used 1.25 mM GTP, all other reactions used 0.3 mM (1/4 of the standard amount).

Amino–G (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
0	46	0
6	54	70
12	50	70
18	56	81
24	10	88

Table: 6.2 Relative total RNA yields and percentages of Amino–G incorporation for each amine concentration used.

Although the use of a reduced GTP concentration results in a drastic loss of total RNA yield, it also provides much improved levels of incorporation. By carrying out transcriptions using Amino–G under these conditions a small amount of RNA may be synthesised, almost 90 % of which is primed with a 5′–amino group.

6.11 Transcription reaction pH measurements

Before beginning this section of research we expected that the addition of sodium hydroxide to the transcription reactions *via* the Amino–G stock could have an adverse effect on the yield of products. To assess the change brought about through the increase in hydroxide concentration we measured the pH change resulting from the use of various quantities of Amino–G stock. These measurements provided an indication to the actual pH of the transcription reaction mixtures, though the products of the reaction were used as the final test to judge whether any adverse processes were taking place. Unfortunately a number of factors made the pH measurements less comparable than we would have liked:

- **Scale:** we were only equipped with a relatively broad tipped pH electrode and therefore required a large volume of transcription reaction mixture to carry out the measurement. Using such a large volume we could not include either the T7 RNAP or RNasin enzymes on the basis of cost. Both enzymes are stored in buffers and, owing to the large quantity of ionisable groups present, are likely to possess large buffering capacity themselves, the addition of these molecules could have a significant effect on the pH of the final reaction mixture.
- **Contamination:** without a surfactant being present proteins such as the T7 RNAP and RNasin enzymes would coat the pH electrode tip, increasing the probability of contamination between solutions.

Table:6.3 below gives the details of the pH measurements:

Amino–G (mM)	pH of reaction mixture	Total RNA yield (%)	5'-AminoRNA (%)
0	7.9	100	0
2.5	8.33	139	4
5	8.72	138	16
7.5	9.54	139	31
15	11.36	102	67
22.5	12.92	80	76

Table:6.3 The data above shows the concentration of Amino–G used in each of the six transcription reaction mixtures, also shown is the pH of each solution. The third and fouth rows give the typical total RNA yields and level of amine incorporation expected when using the related concentration of Amino–G.

We observed the expected increase in pH as the concentration of Amino–G was raised. The pH level measured for the highest concentration of Amino–G is 12.92, well above the optimum level (8) for T7 RNAP. However, as Table 6.3 shows, we have observed RNA transcription using this concentration featuring good total RNA yields and excellent levels of amine incorporation. These two observations can lead us to one of two conclusions:

1. The pH levels we measured were approximately the same as the pH of the final reaction mixtures, and T7 RNAP is more tolerant to high pH levels than previously reported.
2. The inclusion of T7 RNAP and RNasin in the reaction mixture has a significant effect on the buffering capability of the transcription reaction mixture, making the pH levels we measured significantly higher than the actual levels of the full reaction mixture and a poor comparison to the actual system.

6.12 Transcription reaction at pH 8

As the solubility of Amino–G in water is pH dependent, we wanted to investigate the effect that pH had on the products of Amino–G containing transcription reactions. To study the effect of pH on the reaction, and specifically on the solubility of Amino–G we

wanted to undertake a series of transcription reactions using a range of amine concentrations, all at pH 8. We first made up standard transcription reactions containing various amounts of Amino–G stock. The pH of the stock solutions was then lowered to 8 through additions of HCl before the transcription reaction was carried out following the standard protocol (see Section 4.2), recording total RNA yields and the level of Amino–G incorporation for each reaction

6.13 Methodology and results

In the reactions of higher Amino–G concentration this pH change instigated the precipitation of a white solid, from the results of our solubility study (see Section 6.2) we can be confident this solid is Amino–G. Upon addition of radio-label (from a dilute stock to maintain comparability of the transcription reactions) and the T7 RNAP enzyme, transcriptions were carried out following the standard method (Section 4.2). The percentage of 5′-AminoRNA produced was assessed using the sulfo-NHS biotinylation and subsequent PAGE streptavidin assay.

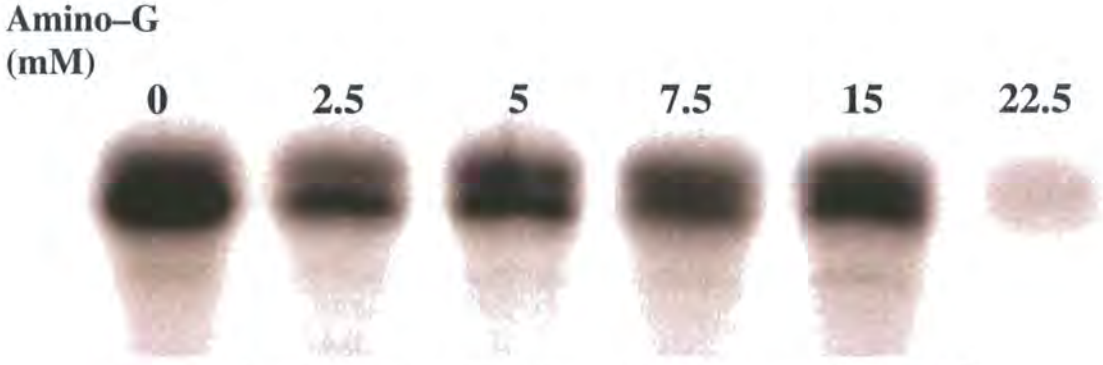


Fig:6.14 IP image of PAGE purification plate. The total yield of ATRib^{TL} transcripts can be seen to vary across the concentration range of amino-G used. So much so that the yield of ATRib^{TL} transcripts was so low in the 22.5 mM reaction that the RNA could not be isolated.

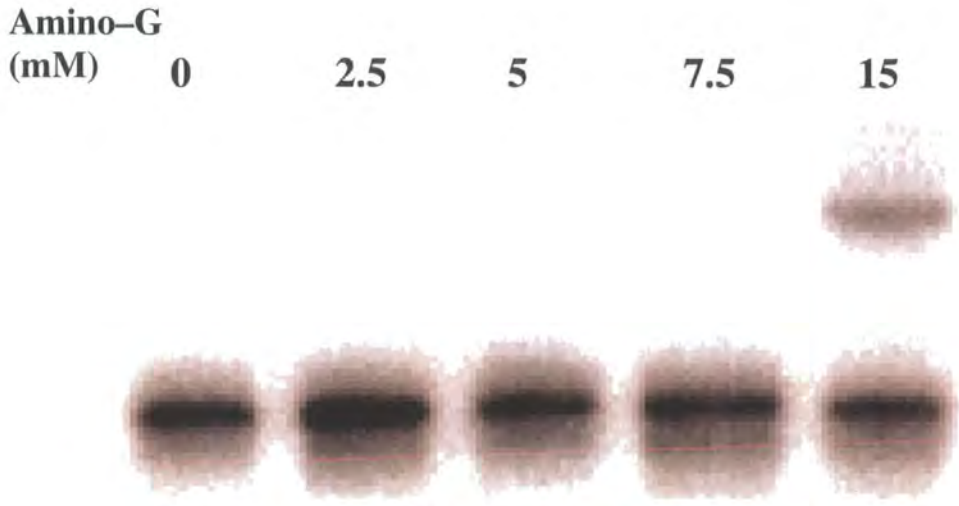


Fig:6.15 IP image of PAGE plate showing results of the streptavidin assay. The image shows separation of two radio labelled RNA transcript populations. Comparison of the two bands intensities allowed the level of Amino–G incorporated into the RNA population to be qualified.

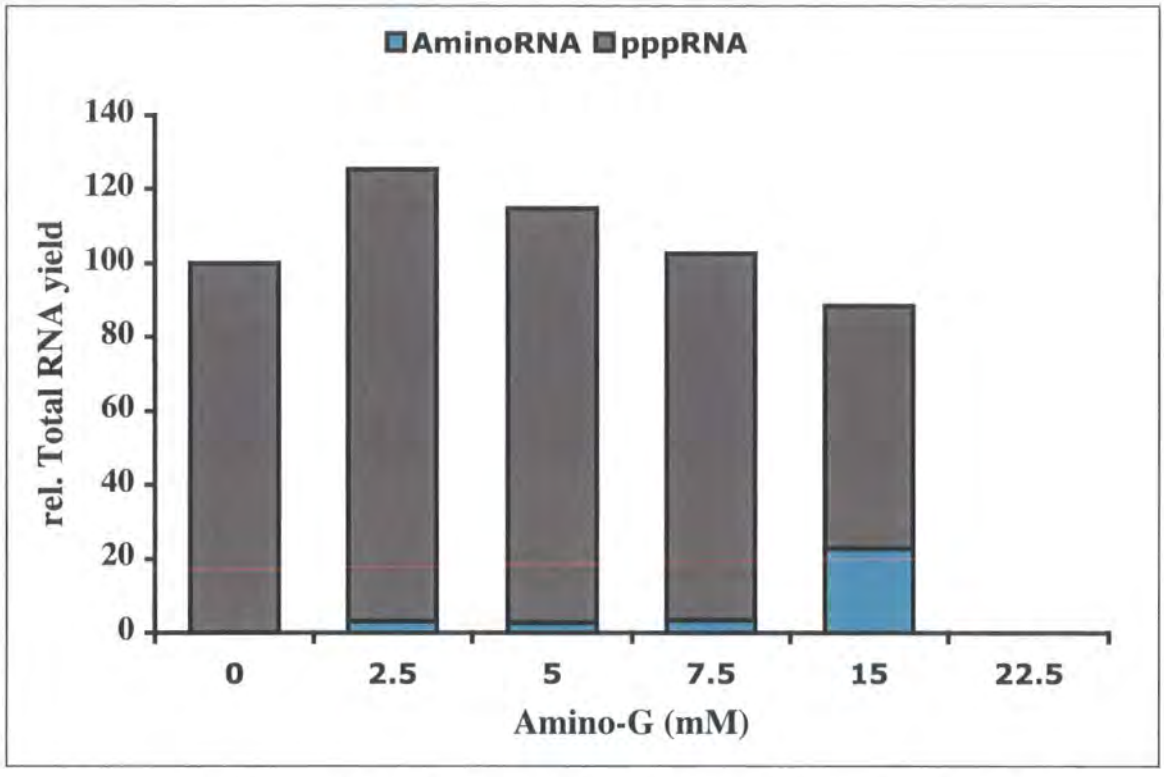


Fig:6.16 Plot of relative total RNA yield *vs* amino-G concentration. The height of each bar corresponds to the relative total RNA yield for each transcription reaction. Each bar is separated into two areas, designating what percentage of the transcripts had been primed with Amino–G to form 5′-AminoRNA.

Amino–G (mM)	Total RNA yield (%)	AminoRNA (%)
0	100	0
2.5	97	2.6
5	98	2.4
7.5	97	3.2
15	74	26
22.5	Trace	–

Table 6.4 Showing the level of incorporation observed when using Amino–G of various concentrations at pH 8.

The experimental data agreed with our prediction that a lowered level of Amino–G incorporation would be observed when the transcription reactions were carried out at low pH. This corroborated our hypothesis that with less amine dissolved in the reaction mixture, competition to initiate RNA transcription sways in favour of GTP, lowering the

incorporation level of Amino–G. As the plot (Fig:6.16, Table:6.4) demonstrates there was a massive reduction in the amount of 5′–AminoRNA produced in comparison to previous experiments using equivalent amounts of Amino–G, but without pH re-adjustment (see Table 6.3). Total RNA yield did not suffer from the addition of the novel nucleoside except in the highest Amino–G concentration studied. The highest level of incorporation was 25 %, close to the maximum level previously quoted in literature.

6.14 Amino–G Conclusions

Our targets when beginning this area of research were:

- To establish whether the use of Amino–G solubilised in a weak $NaOH_{(aq)}$ solution could prime T7 RNA polymerase mediated transcription reactions with more success than previous methodologies have reported.
- To then optimise the amount of Amino–G in the transcription reaction to maximise 5′–AminoRNA production, taking into account both the total RNA yield and level of amine incorporation.
- To determine what effect, if any, the hydroxide in the amine stock has upon the total RNA yield and the level of Amino–G incorporation.

Of these aims the first two were the most straightforward and easily met. Before addressing the results of our pH/ solubility studies I will first summarize the various product distributions observed from Amino–G incorporating transcription reactions.

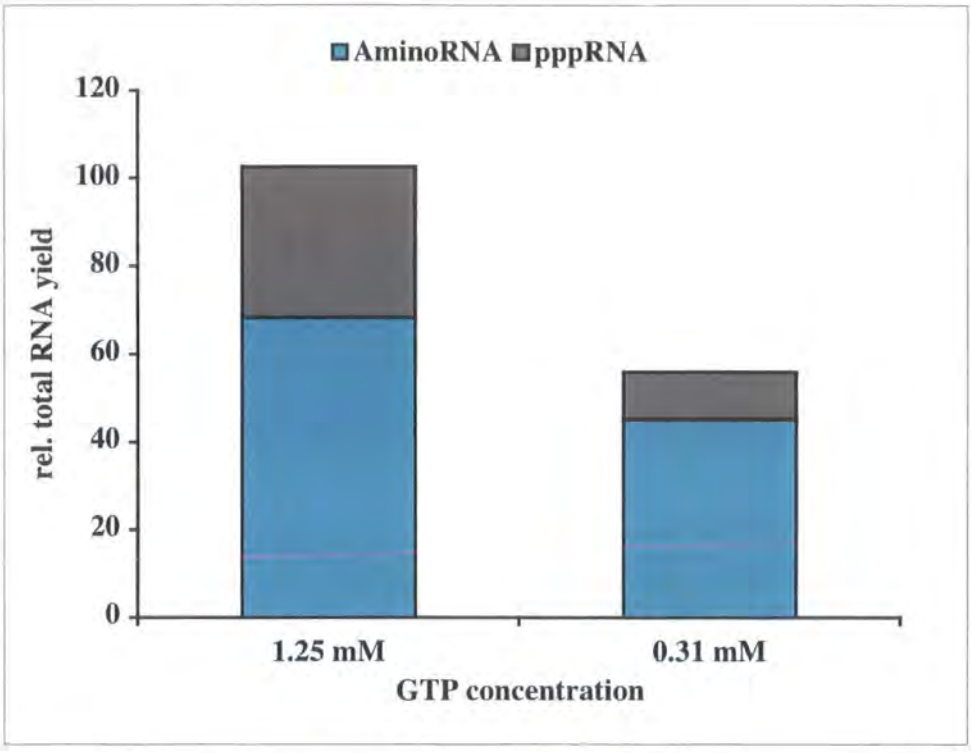


Fig: 6.17 Plot of RNA transcript products observed when using standard (1.25 mM) and reduced (0.31 mM) GTP concentrations in Amino–G incorporating transcription reactions. Results shown are those of the optimum for each experiment, showing a balance of total RNA yield and amine incorporation.

In fitting with our expectations we observed the highest total RNA yield when using the standard (1.25 mM) amount of GTP in transcription reactions. In reducing the GTP concentration we noted an increase in the level of Amino–G incorporation, from 76 – 88 %, however the low concentration of GTP present resulted in a reduced total RNA yield. As with GANP we can envision situations in which the use of either methodology (standard or reduced GTP concentration) could prove valuable.

1. If the requirement for the transcription reaction is to produce RNA transcripts the maximum amount of which incorporates an amine group at the 5′–terminus, such as in the production of a randomly sequenced RNA library for selection studies, then the use of a reduced GTP concentration is endorsed.
2. If a larger quantity of 5′–AminoRNA were required then it would be possible to increase the scale of the reduced GTP reaction. However, it may prove more

practical and convenient to use the standard methodology and sacrifice a very high level of incorporation for an increased overall RNA yield.

The level of Amino–G incorporation observed for both methodologies is far higher than that quoted in the literature. Our research has provided evidence supporting the hypothesis that the improvement in Amino–G activity can be attributed to its increased availability in the reaction mixture *i.e.* the revelation of the amine's pH dependent solubility in aqueous solutions:

- Our study using UV spectroscopy (Section 6.2) to illustrate the pH dependent solubility of Amino–G showed a marked difference between the concentrations attainable in pH 13 and pH 9 solutions. From this we can infer that more amine was available transcription reactions using Amino–G pre-dissolved in $NaOH_{(aq)}$ (pH 13) than when following the methodology of Suga *et al* (pH 9, very little amine in solution), resulting in higher levels of incorporation in comparable reactions (76% and 20% respectively)[45].
- When using pre-dissolved Amino–G in transcription reactions adjusted to pH 8 (Section 6.12) the maximum level of incorporation observed was close to that reported by Suga *et al* (26%). We can attribute the decrease in level of incorporation to the precipitation of Amino–G from the reaction mixture as the pH of the solution was lowered, as was demonstrated in the UV spectroscopy study.

Using these observations we have proposed that in transcription reactions carried out using a pre–dissolved stock of Amino–G either:

1. The pH levels of the reactions were similar to those used by Suga *et al*, but as the Amino–G was pre-dissolved in our experiments, more remained in solution as the pH was lowered than would be present if the amine were added as a solid.
2. The reactions took place at a pH higher than those of Suga *et al*, and therefore contained a higher concentration of the amine.

We have collected evidence to support both of these mutually exclusive propositions. Our UV spectroscopy study adequately illustrated that when reducing the pH of

Amino–G solutions from 13 to 9 some amine was lost as precipitate, although the resulting Amino–G concentrations were still higher than that of the pH 9 buffered solutions used by Suga *et al* (Section 6.2). The difference in Amino–G concentration in solutions of the same pH illustrates how the route to making a solution can be as important as what the solution contains. This evidence provides strong support for the first of the two hypotheses proposed above, however the data is flawed in that the UV spectroscopy and pH measurements were made on aqueous solutions containing only Amino–G and NaOH and therefore not directly analogous to the conditions present in the transcription reaction mixture. We believe that the analogous behaviour would be observed under transcription reaction conditions, as there would almost certainly be a lowering in pH from that of the Amino–G stock solution. However as we were unable to make measurements of the pH change that takes place in the transcription reaction mixture upon Amino–G addition we cannot make a definite judgement. This leads us on to the second of the two proposed theories for Amino–G's increased levels of incorporation; the transcription reaction taking place at a pH above 9.

Although we were unable to measure the pH of the complete Amino–G containing transcription reaction mixtures, we were able to measure the pH prior to the addition of the T7 RNAP and RNasin enzymes (Section 6.11). As Table 6.3 shows we recorded pH levels higher than 8, the optimum for T7 RNAP mediated transcription reactions.

Amino–G (mM)	**pH of reaction mixture**	**Total RNA yield (%)**	**5'-AminoRNA (%)**
0	7.9	100	0
2.5	8.33	139	4
5	8.72	138	16
7.5	9.54	139	31
15	11.36	102	67
22.5	12.92	80	76

Table:6.3 The data above shows the concentration of Amino–G used in each of the six transcription reaction mixtures, also shown is the pH of each solution. The third and fouth columns give the typical total RNA yields and level of amine incorporation expected when using the related concentration of Amino–G.

These measurements would seem to support the proposition that the high levels of Amino–G incorporation we observed were owing to an increased concentration of the amine in the reaction mixture achievable because of the high pH of the solution. However, this is at odds with literature data, which states that not only is RNA readily hydrolysed in a matter of minutes at elevated pH levels but that the amount of RNA transcripts produced in a transcription reaction is reduced when the reaction mixture deviates from the optimum value (20% for +/_ 1 pH level)[2,63]. As Table 6.3 shows, the recorded pH levels and total RNA yields do not follow this trend, indicating that either our measurements or those in the literature are incorrect. There is the possibility that the T7 RNAP stock we prepared was more active than that used in other workers research, resulting in sustained activity in higher pH environments, however this is merely conjecture with no evidence to support the claim. Unfortunately owing to limitations placed on our ability to make the pH measurements, we have reason to believe our results are not as comparable to the true reaction mixture conditions as they could have been.

For practical reasons the pH measurements of the reaction mixtures were made in the absence of both T7 RNAP and RNasin and therefore lacked both the buffering capability of the enzymes and the TRIS buffers present in the enzyme storage solutions (Section 6.11). We cannot disregard the role these species play in controlling the pH of the transcription reaction mixtures; however neither did we have the practical requirements to accurately assess their buffering capacity. This uncertainty creates adequate error in our measurements to doubt their applicability to 'true' Amino–G incorporating transcription reaction conditions. Hence, removing supporting evidence for the proposition that the high level of amine incorporation observed in our study was owing to a heightened pH in the transcription reaction.

To verify our pH measurements a systematic study would be required, looking at complete transcription reaction mixtures incorporating a range of Amino–G concentrations. By ensuring all components of the reaction are present the pH measurements made would be directly applicable to the products of any transcription

reactions undertaken, from which greater understanding of Amino–G's behaviour in the reaction could be gained. By running transcription reactions under precise pH control we would also be able to assess the relative activity of the T7 RNA polymerase under unfavourable conditions. This information would aid in raising our understanding of how the reaction proceeds under high pH conditions, and could allow us to rule out degradation of the enzyme as a reason to limit the pH of Amino–G incorporating transcription reactions.

The key requirement to making these measurements would be the availability of a very fine tipped pH electrode, capable of analyzing solutions of low volumes *e.g.* 50–100 μL. As we possessed neither a suitable electrode nor the time required to undertake the research we could not collect this data ourselves. Based on the incompatible data for the decrease in RNA yield with rising pH and the total RNA yield levels observed in our own research, we hypothesise that measurements made of complete Amino–G incorporating transcription reactions would reveal pH values closer to the optimum of 8 than our own measurements.

To reiterate, our two hypotheses for Amino–G's relatively high level of incorporation were:

1. The pH levels of the reactions were similar to those used by Suga *et al*, but as the Amino–G was pre-dissolved in our experiments, more remained in solution as the pH was lowered than would be present if the amine was added as a solid.
2. The reaction took place at a pH higher than those of Suga *et al*, and therefore contained a higher concentration of the amine.

The results from our UV spectroscopy study corroborate the first of these propositions, while the second seems extremely unlikely owing to its contradiction of literature observations. Therefore, our opinion is weighted towards the dissolution of the amine in a basic solution prior to use in transcription reactions is key to its efficient use, although

to make a definitive judgement more evidence regarding the actual pH of Amino–G incorporating transcription reactions would be required.

7.0 Azido–G Transcription Study

7.1 Reasoning for Azido–G study

The dissolution of Amino–G in a basic solution for use in transcription reactions provided a route for the use of other 'insoluble' guanosine derivatives. We chose to study the azide precursor to Amino–G as not only was the synthetic route already in place, but the incorporation of an azide group into the 5′–terminus of RNA molecules could enable us to undertake some interesting chemistry. Azido–G displays solubility similar to that of the amine, as we observed a level of azide incorporation we can presume the same key factors also appear to govern its ability to initiate transcription. Through the use of phosphine chemistry we were able to reduce the incorporated terminal azide *in situ* to form 5′-AminoRNA. As in our previous studies this amine group could then be used as a bioconjugative 'handle'[64,65].

In addition to opening a third route to 5'–AminoRNA, the incorporation of a 5′–terminal azide group also presented the possibility of undertaking a Copper(I) Catalyzed Azide–Alkyne Cycloaddition (CuAAC) reaction to bioconjugate to the RNA transcripts[64,66]. This metal ion mediated coupling methodology is excellent example of a "click chemistry" process and has been applied within biological systems without resulting in unwanted side reactions[64]. We observed a successful bioconjugation, however did not attempt to optimise the reaction and therefore were unable to carry out a systematic study regarding the efficiency of the coupling.

Literature examples exist wherein azido-terminated RNA has been used *via* cross-linking reactions to study the binding sites of enzymes, however synthesis of this azide terminated RNA took place over a number of steps involving bioconjugation to GMPS primed RNA[30]. The synthesis of 5′–AzidoRNA in good yield over one step, using a novel nucleotide that is more convenient to synthesise than GMPS, would be an obvious improvement to the current methodology.

7.2 Targets for Azido–G

Our targets for the use of Azido–G were less extensive than those previously set for research into incorporation of Amino–G and GANP:

- Demonstrate the ability for Azido–G to initiate T7 RNAP mediated transcription reactions to form 5′–AzidoRNA.
- Demonstrate that the azide group of 5′–AzidoRNA can be reduced *in situ* to give 5′–AminoRNA.
- Investigate whether the azide group of 5′–AzidoRNA could take part in a CuAAC bioconjugation reaction.

7.3 Transcription Reactions with Azido–G

Solubility of Azido–G was comparable to that of Amino-G, *i.e.* when attempting to make a 100 mM stock of Azido-G in neutral pH water no dissolution was observed, however, upon addition of base to make a 100 mM NaOH solution the solid completely dissolved to produce a clear, slightly yellow solution. The hypothesis of N^6 amine deprotonation used to explain Amino–G's pH dependent solubility could also be applied to Azido–G owing to its almost identical structure. The similar solubility behaviour allowed us to study Azido–G *via* transcription reactions in a methodology almost identical to our Amino–G and GANP investigations.

Through the use of various concentrations of Azido–G in a series of transcription reactions we carried out an investigation to determine the highest 5′-AzidoRNA yield possible per reaction. The transcription reaction methodology followed was identical to that used for Amino-G and GANP. The use of Zeba Desalt spin columns was not necessary for the purification of the RNA produced, as the residual Azido–G did not run coincident with the transcripts.

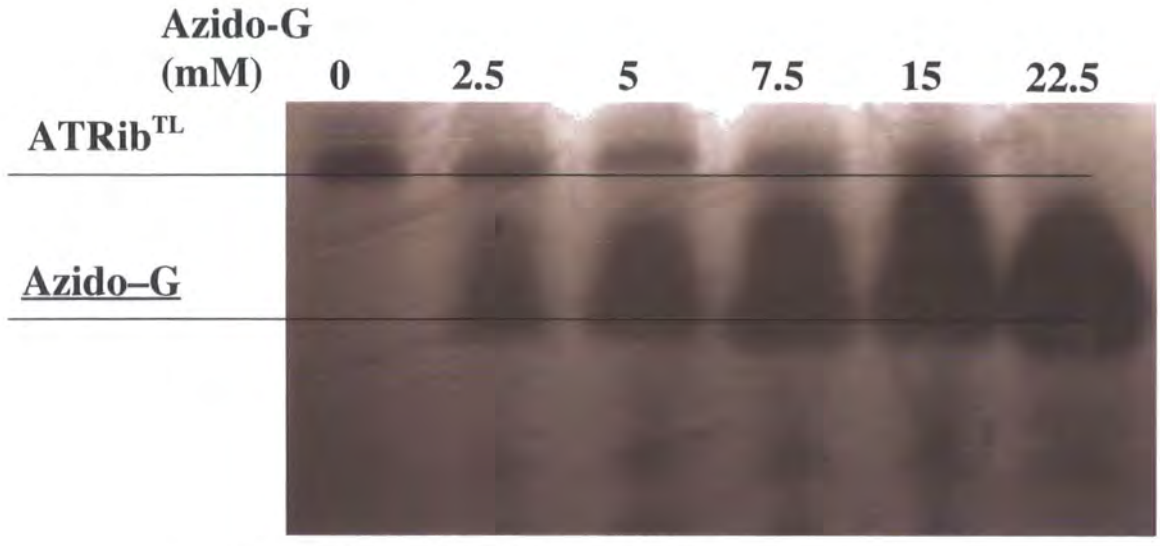


Fig:7.1 UV shadowing image of PAGE gel used to purify transcription reactions of varying Azido–G concentration.

Following incubation for 2.5 h each transcription reaction mixture was diluted in urea loading buffer, heated to 95 °C for 1 min, and then loaded on to the PAGE gel. RNA bands were excised from the gel and eluted into $NaCl_{(aq)}$ solution before removal of the residual gel and precipitating the RNA through addition of ethanol as described in the transcription chapter (see Section 4.5). The transcripts were then assessed for their levels of azide incorporation *via* two methods; a reduction method, and a 'click' method.

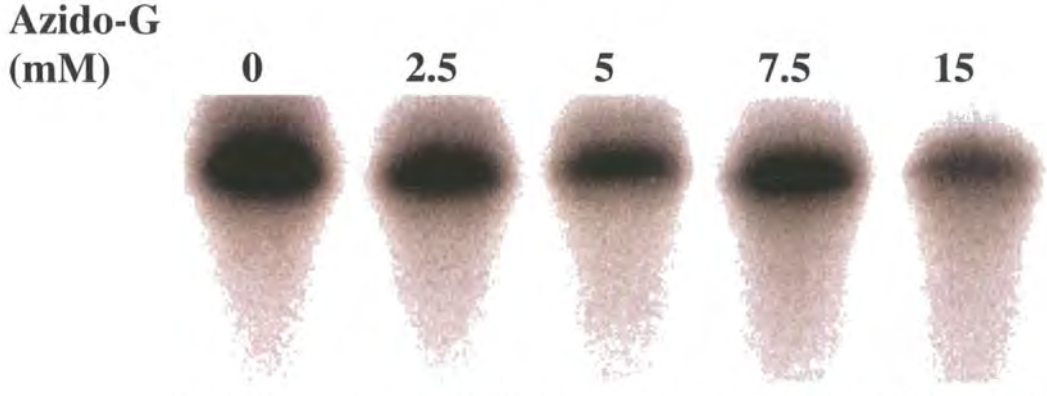


Fig:7.2 IP image of a PAGE gel showing purification of radiolabelled 5′–AzidoRNA/ RNA transcripts.

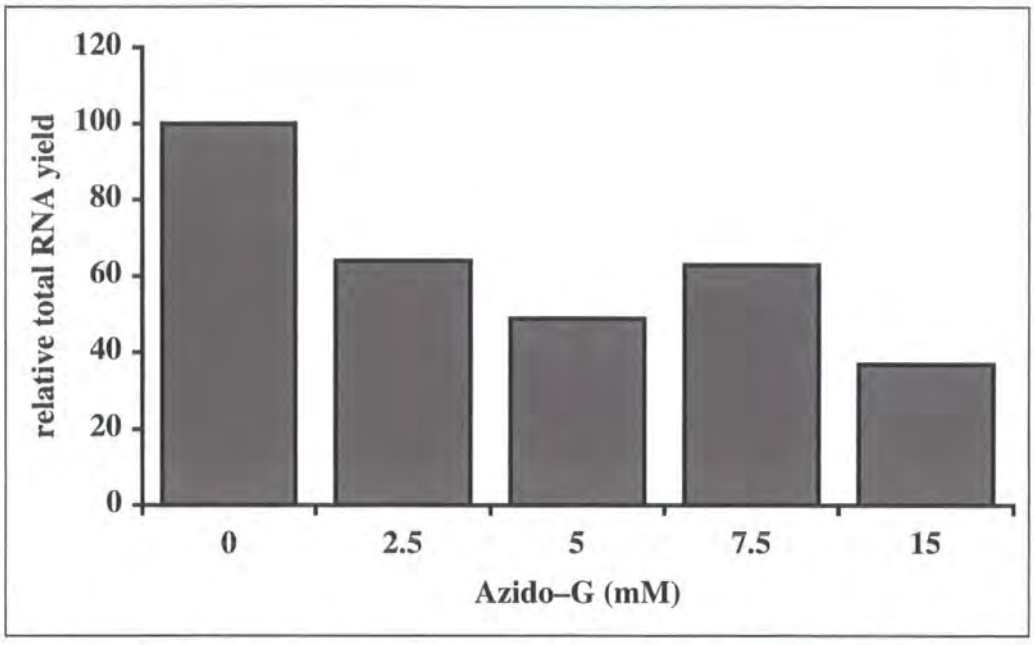


Fig:7.3 Plot showing the relative total RNA yield of a transcription reaction *vs* the concentration of Azido–G in the reaction.

7.4 *In stitu* Azide reduction

A number of reducing agents for biological materials are available and in common use. The water soluble tris(2-carboxyethyl)phosphine hydrochloride (TCEP) reducing agent was deemed sufficient for our needs. This phosphine is commonly used for the reduction of disulfide bonds, however it has been shown to be powerful enough to reduce azide groups *via* a mechanism similar to that of the Staudinger reduction.

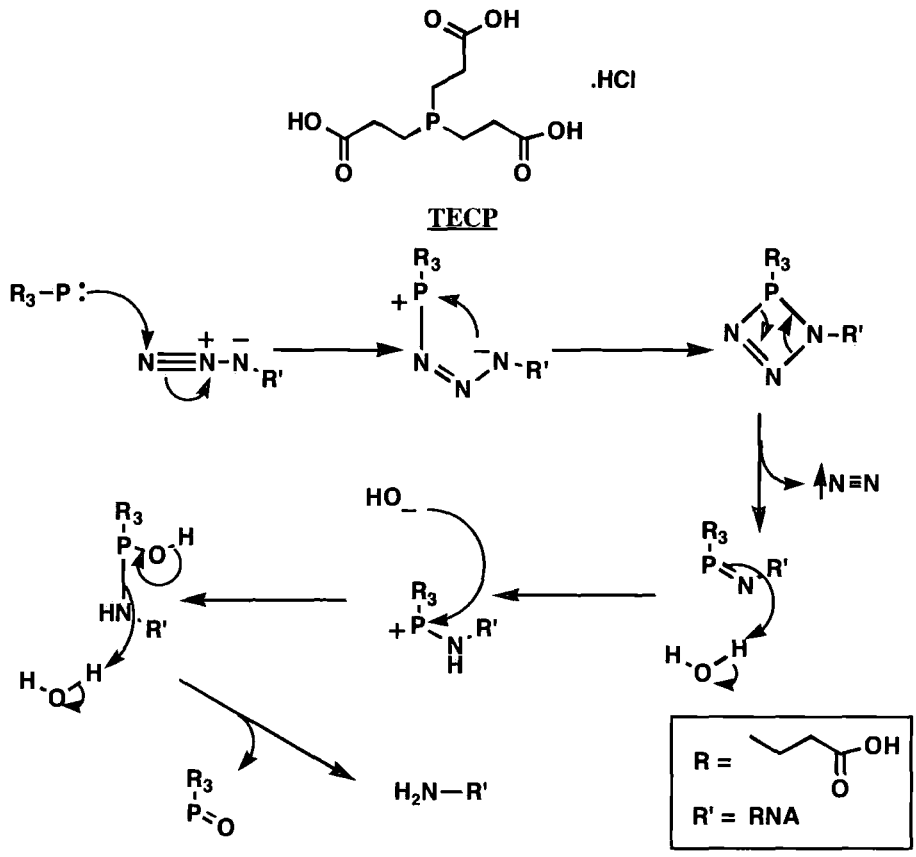


Fig:7.4 Proposed reaction mechanism for the reduction of the terminal azide of Azido–RNA by TCEP.

A sample of RNA from each transcription reaction was incubated with TCEP in HEPES buffer (pH 8) for 1 h to reduce any azide groups present. The resulting amine groups were then biotinylated by the same procedure used in our GANP and Amino-G investigations through the use of sulfo-NHS-biotin. As in previous our studies the use of gel electrophoresis in conjunction with streptavidin conjugation was used to assess the level of Azido–G incorporation (Section 4.10). As the biotinylation was dependent upon the success of the TCEP reduction, and we had no way of assessing the yield of this reaction at the scale we used, we assumed all azide groups present were reduced and biotinylated. Therefore the results quoted are not exact amounts, but minimum levels of incorporation.

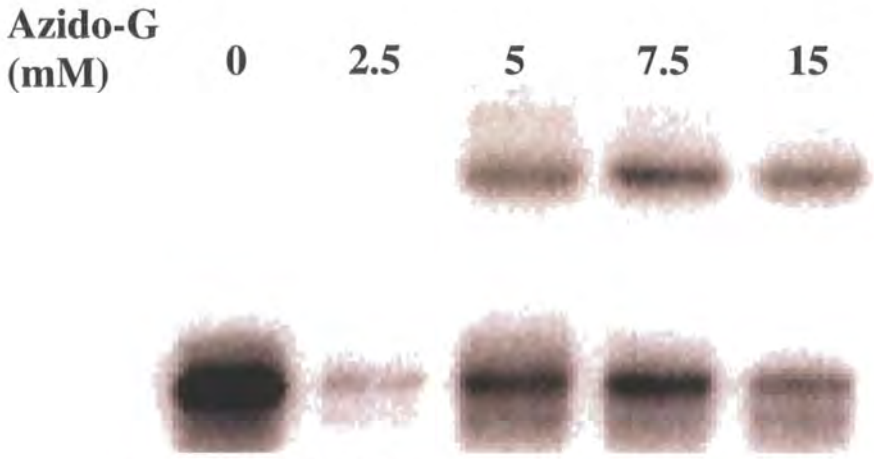


Fig7.5 IP Image of PAGE gel showing results of a Streptavidin assay using TCEP reduced and amine–specifically biotinylated Azido–G containing transcription reaction products.

We ran a control gel to confirm the retarded bands are owing to biotinylated RNA-streptavidn conjugates, and not interactions between streptavidin and non-biotinylated azido/aminoRNA.

Azide	+	+	–	–	+	+
Biotin	+	+	+	+	–	–
TCEP	+	+	+	+	+	+
SaV	+	–	+	–	+	–

Azide	–	–	+	+	–	–
Biotin	–	–	–	–	–	–
TCEP	+	+	–	–	–	–
SaV	+	–	+	–	+	–

Fig:7.6 IP image of PAGE gels showing control experiments demonstrating the selectivity for streptavidin binding with Azido–G primed RNA that had been reduced with TCEP and subsequently biotinylated. The experiments show a retarded band in the PAGE gel can occur through no other combination of RNA species and streptavidin.

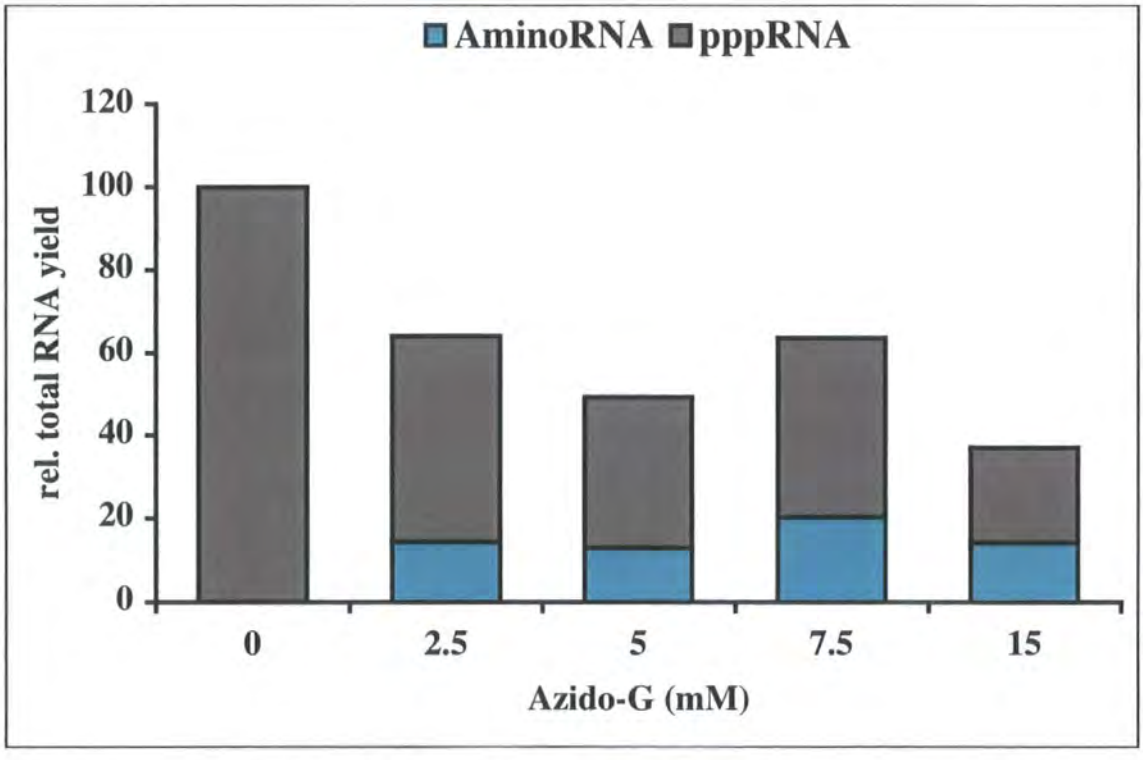


Fig:7.7 Plot of relative total RNA yield (w.r.t the 0 mM control) vs Azido-G concentration.

Azido-G (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
2.5	64	23
5	49	26
7.5	63	31
15	37	38

Table:7.8 Showing the data used to produce Fig:7.7.

Although we have demonstrated that Azido–G can initiate RNA transcription and that an *in situ* reduction to the amine can be achieved, the efficiency and yield are not as high as those achieved with Amino–G or GANP. If the synthetic step between the azide and amine was complex and low yielding, there could be an argument for the use of Azido–G. as a route to 5′–AminoRNA. However, as the reduction used to convert the azide to Amino–G, is direct, high yielding, and not synthetically challenging, there seems little point in using the azide in transcription reactions for the synthesis of 5′-AminoRNA.

With the solubility of Azido–G being comparable to the more efficiently incorporated Amino–G, the low total RNA yields and levels of incorporation are likely to be owing to unfavourable interactions occurring between the initiator and T7 enzyme. The exact nature of these interactions is difficult to assign given the complex nature of enzyme active sites.

7.5 Copper catalyzed conjugation to the azide

We could find no existing protocols for the direct bioconjugation to a 5′–AzidoRNA molecule; therefore we thought to devise a method, based on other CuAAC reactions involving biological materials[30]. Our first step was to acquire or synthesise a suitable alkyne containing reporter group that could be conjugated to the terminal azide. As we had successfully used streptavidin conjugation assays in previous studies we chose to use a similar analytical method in this research. We were able to obtain (courtesy of Prof. R. Edwards) a biotinylation reagent in which the biotin structure is linked to a terminal alkyne group (see Fig:7.8). The transformation we intended to undertake is detailed below:

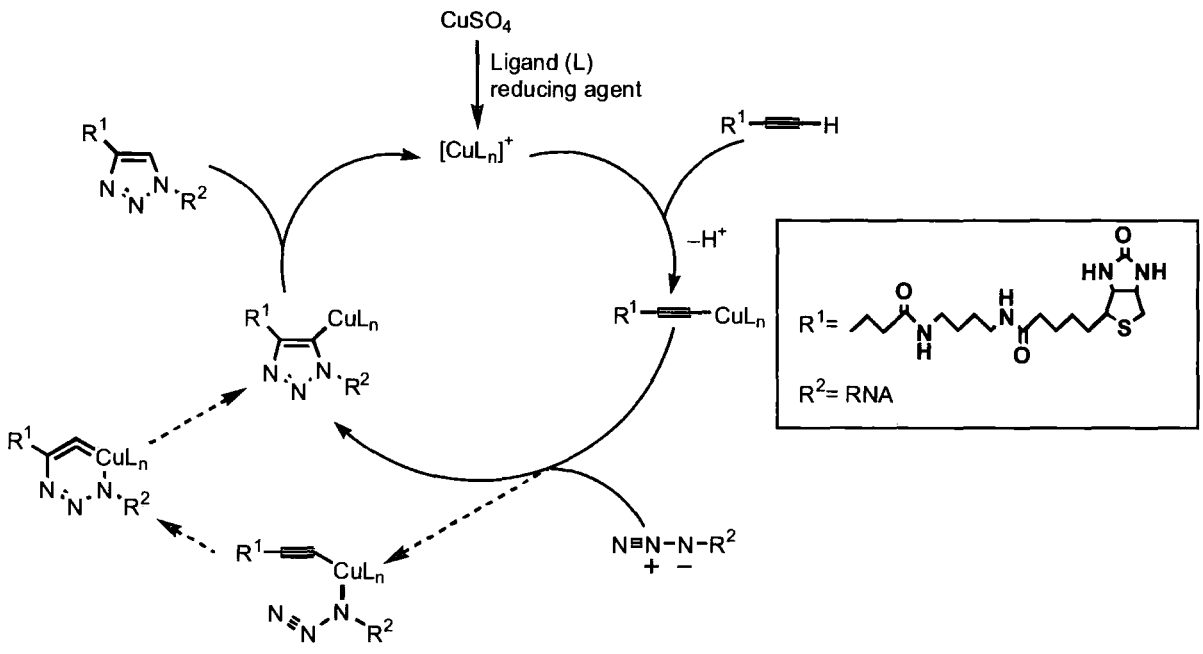


Fig:7.8 Proposed scheme for a Copper Catalyzed Azide–Alkyne Cycloaddition reaction[66].

Rather than starting from first principles we chose to mimic our CuAAC reaction conditions on reported in the literature, then if we had any success the method could be optimised at a later date. The conditions we chose to use were those reported for the coupling of azide functionalized PEG spacers to 5-hexynoic acid[30].

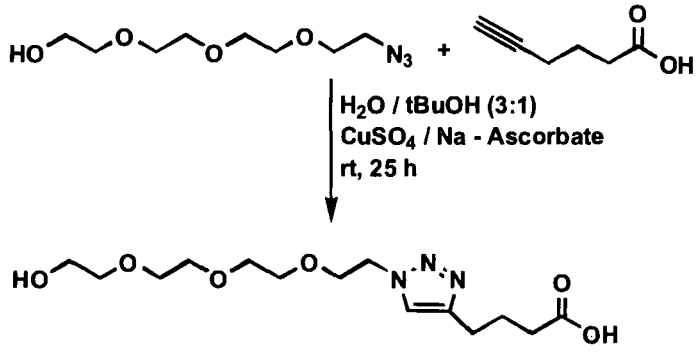


Fig:7.9 Reagents and conditions for a CuAAC reaction used to couple azide terminated PEG molecules to an alkyne terminated acid[30].

The reaction scheme occurs *via* the *in situ* generation of a Cu(I) reaction catalyst *via* reduction of Cu(II) from $CuSO_4$ by sodium ascorbate. In developing this reaction the authors observed none of the desired product in the absence of catalyst. The procedure was found to be highly

reliable, giving yields in excess of 90%, and therefore provided a good basis for developing our own bioconjugation methodology.

7.6 First attempt at CuAAC assay

For our first attempts at a CuAAC bioconjugation to 5′–AzidoRNA we followed the conditions used in the literature model reaction exactly. The azide substrate was RNA purified from the transcription reactions carried out in the presence of Azido–G as described earlier (see Section 7.3).

After the 25 h reaction time there was a solid visible in each reaction mixture, the brown/yellow colour of which lead us to believe it was likely to be a copper salt. Upon attempting to remove the precipitate by centrifuge we found that ~90 % of the radio–labelled material present in the reaction mixture (as assessed by Giger meter) had also precipitated from the solution. We attempted to re-dissolve the transcripts from the precipitate pellet in DEPC water; unfortunately this also resulted in the re–suspension of the residual solid. Subsequent attempts to redissolved the RNA in $NaCl_{(aq)}$ solutions could not separate our desired product from the solid material. Therefore, as a method of analysing the preliminary reaction we undertook PAGE of the crude product, including the suspended solids. Two lanes were used, with half the reaction mixture in each, one lane used a streptavidin/ urea loading–buffer, while the second used a urea only buffer.

Fig:7.10 IP image of a PAGE gel showing the results of a streptavidin assay. The two lanes were both loaded with 5′–Azido–RNA that had been subjected to a CuAAC bioconjugation reaction, only one lane however also contained streptavidin in the buffer.

In the streptavidin–containing lane a proportion of the RNA was retarded, indicating biotinylation was a success. However, this is the only clearly resolved band in the image, the other radio–label signals were both streaked and poorly resolved. These poor signals are likely to be the result of the presence of solids and excess salt in the crude reaction mixtures loaded to the gel.

7.7 Results of CuAAC assay

From our first set of results we were been able to ascertain that bioconjugation was successful, meaning the incorporated azide is available for CuAAC reactions. Unfortunately we cannot determine the level of incorporation *via* the results collected thus far as the resolution of the PAGE gel is far too poor to assess the relative amount of RNA present in each band. In an attempt to improve the quality of the assay we re-

assessed the bioconjugation reaction, believing that the solid present at the end of the reaction, was the main source of poor band resolution on the PAGE gel.

7.8 Modifications to CuAAC method

In an effort to minimise the quantity of solid residue in the reaction mixture we reduced the concentration of metal ions and ascorbate present in the mixture. We believed that having less metal salts to remove should allow us to produce a clearer PAGE image, the downside being we could potentially generate an inadequate amount of Cu(I) catalyst to complete the conjugation reaction. Upon reviewing the levels of Azido–G incorporation as determined by the reducing method, we can see the maximum percentage of incorporation is less than 40 %. Therefore as we had judged our initial concentrations of $CuSO_4$ and sodium ascorbate relative to the total amount of RNA in the bioconjugation reaction, we felt a reduction in the amount of reagents could be made without affecting the resulting yield of coupled product.

Unfortunately, although we were able to reduce the amount of solid impurities present in the crude reaction product, we were unable to satisfactorily clean the samples prior to PAGE. In attempting to remove the solid, whether it was by centrifuge or filtration, we could not avoid also losing RNA product. Some improvement in the quality of the acquired results was achieved (as can be seen in Fig:7.11). However the change was not significant enough to allow us to qualify the percentage of bioconjugation that took place. This was unfortunate, as we would have liked to have–made comparisons between the direct coupling and reduction methods of bioconjugation.

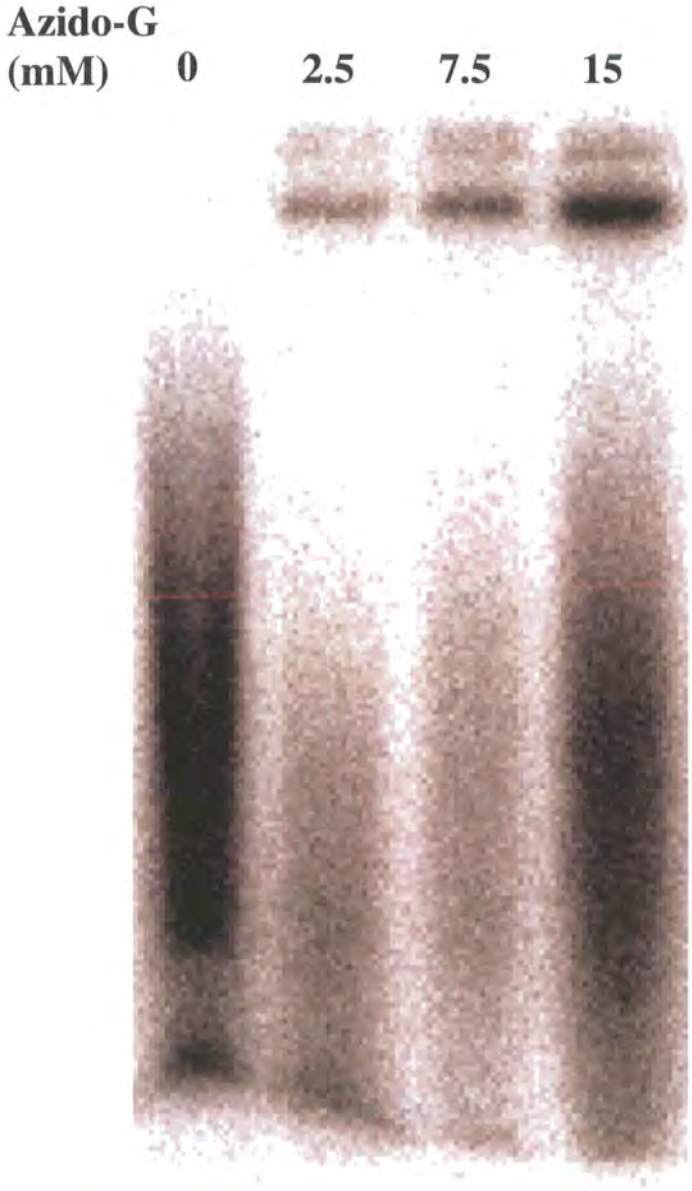


Fig:7.11 IP image of a PAGE gel showing separation of CuAAC coupling reactions. The substrates for these reactions had been transcribed in the presence of various amounts of Azido–G.

7.9 Azido–G Conclusions

Owing to time constraints we were forced to call a halt our studies into the direct conjugation to 5′–AzidoRNA. Although we were unable to develop a method for the clean bioconjugation to 5′-AzidoRNA *via* a CuAAC method, we did ascertain that such a reaction is possible. Our initial reaction and resulting +/– streptavidin assay confirms that the azide group is available for reaction and can act as a 'handle' for bioconjugation.

The poor resolution observed in all the assay gels ran in this series of experiments could have been owing to the presence of residual solids in the reaction mixtures. Our inability to remove this solid without also losing the majority of our RNA product was a major factor in determining the effect the solid was having our results. We believe that ionic binding occurring between the anionic phosphate backbone of the RNA and the cationic metal ions could have been the source of the difficulty we encountered in separating the two species. Removal of metal ions could be achieved through the use of multi-dentate ligands such as EDTA.

The source of the poor band resolution in the PAGE gels could also have been owing to the presence of hydrolysed RNA transcripts in the reaction mixture. As the reaction mixture contains metal ions it is possible that hydrolysis of the phosphoester backbone had taken place. This would result in a population of RNA transcripts of various lengths and radioactivity. When separated by PAGE using a gel of moderate cross-linking such a population of short transcripts would not be well resolved, as in our IP images. The mass difference between a full ATRib^{TL} transcript (~10 kDa) and the streptavidin protein (~60 kDa) is such that all biotinylated transcripts, regardless of their length, would likely run together owing to the small contribution the RNA makes to the total mass of the conjugate. Unfortunately the amount of time we had to investigate the coupling reaction was insufficient to study these possibilities further.

Investing further time in optimizing the biotinylation reaction seemed ill advised when taking into account we posses another method of assessing levels of Azido–G incorporation *via* reduction with TCEP and subsequent biotinylation. It would be more reasonable that a research group requiring to couple to 5′-AzidoRNA develop a methodology specific to their own needs now we have confirmed the methodology has promise.

8.0 Summary of Results

8.1 Aims

This chapter will act as a summary of all the conclusions made during the study of GANP, Amino–G and Azido–G. We will begin by comparing the various total RNA yields and levels of incorporation achieved using each of the nucleotides and proposing reasons for the observed results. The success of the study will then be assessed and some areas of future work discussed.

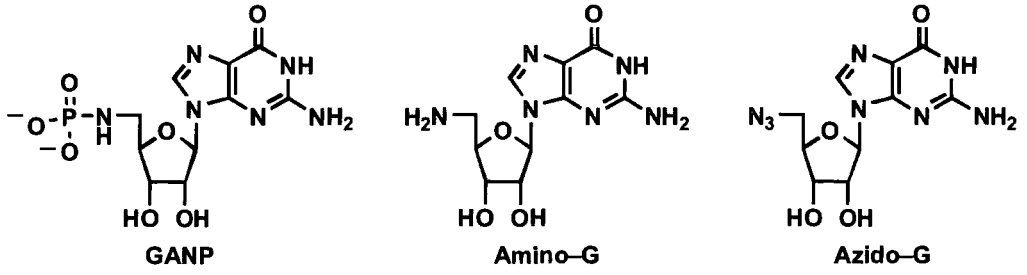


Fig:8.1 The novel nucleotides, and nucleosides we used to initiate transcription reactions and produce 5′–AzaRNA.

8.2 Comparative Results

As all of the transcription reactions we have undertaken during our research, whether they incorporated GANP, Amino–G, or Azido–G, have followed a standard methodology the results of these experiments are comparable with each other. This also holds true for the methods used to determine the level of novel nucleotide incorporation as they were all based on the same biotin–streptavidin interaction and were assessed using IP exposure technology. The exact nature of the biotinylation reactions applied to each species of 5′–AzaRNA did vary, though in each case the reagent was used in great excess, with the reaction assumed to have achieved 100 % conversion. This assumption means that all values of incorporation levels we have quoted are in fact minimums, allowing for the possibility of a small amount of unreacted 5′–AzaRNA being present and undetected in any given RNA population.

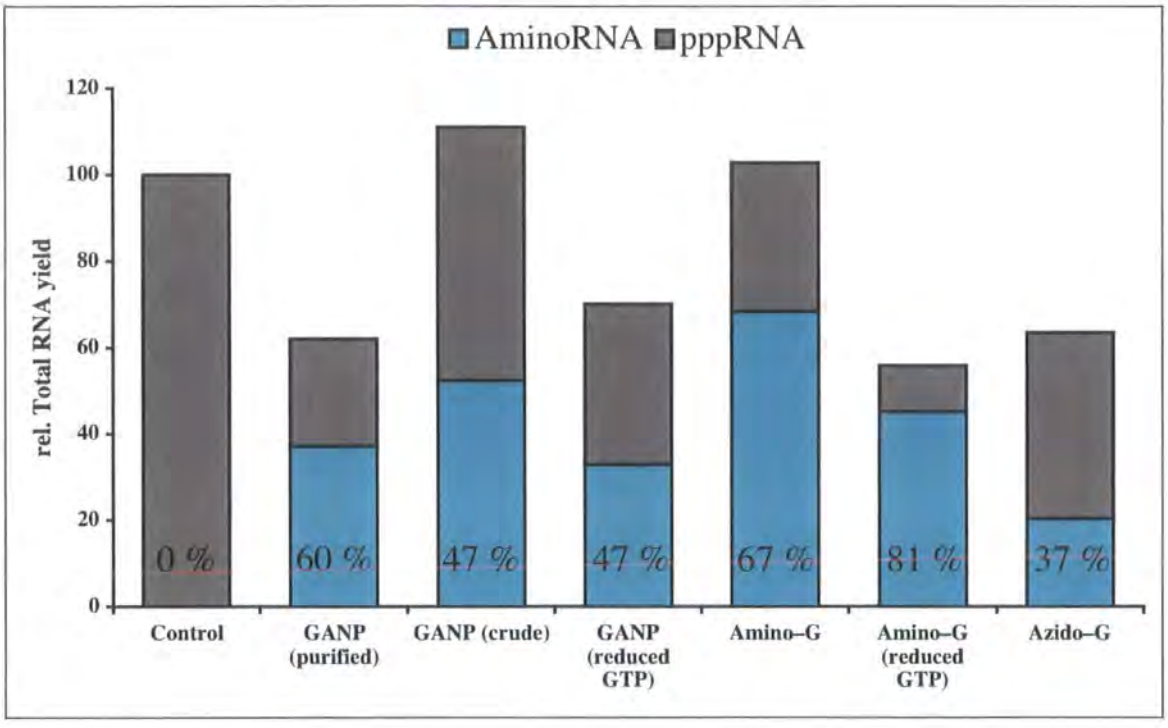


Fig:8.2 Plot of relative Total RNA yield *vs* the range of 5′–Aza–Guanosine nucleotides studied. Factors affecting the observed results are given in parentheses. The level of incorporation of each novel nucleotide is superimposed on the AminoRNA section of each column.

The target of our research was to devise a simple and reliable route to the efficient synthesis of 5′–AminoRNA. In terms of the levels of incorporation achieved it is clear that improvements have been made on the 20 % reported by Suga *et al* for their use of Amino–G[45]. As has been previously discussed this improvement has been attributed to the increased solubility of the novel nucleotides in aqueous solutions, whether through the addition of an ionisable phosphate group in the case of GANP or the exploitation of the pH dependent solubilities of Amino–G and Azido–G.

8.3 GANP Summary

Focussing on the results for GANP we can see that our attachment of a masking group to Amino–G successfully increased the solubility of the nucleotide, allowing it to be dissolved in a neutral aqueous solution and used in transcription reactions. The products of these reactions, whether using a crude or purified sample of the phosphoramidate,

displayed levels of incorporation more than double that reported by Suga *et al* when using Amino–G[45]. Had our studies halted at this point we would still have produced a successful result and improved on the routes to 5′–AminoRNA already available. As covered earlier (Section 1.32) previously reported methods either resulted in the poor levels of incorporation (Suga *et al*) or involved complex and unattractive synthetic methods (Amino–PEG)[31,45]. Although the purification of GANP following the aqueous amino–phosphorylation reaction is a time consuming process the other steps in the synthesis are relatively simple, for example the work–up procedures avoid the use of any chromatographic methods. Also in our favour is the fact that it is not necessary to purify GANP from the residual Amino–G and sodium salts present in the crude reaction mixture. As Fig:8.2 shows the total RNA yields and level of incorporation achieved when using crude GANP are more than adequate to argue forgoing the purification entirely and accepting the reduced level of novel nucleotide incorporation. As with all things the choice of whether to purify the GANP would rest on the individual user, who would have to balance their own needs for achieving a higher level of incorporation with the time taken to complete the purification.

As mentioned above, had our research been halted having only studied GANP the route to 5′–AminoRNA developed would have met all of our targets and we would have deemed the work a success. However, during the synthesis of GANP we had observed the increased solubility of Amino–G in basic solutions, compared to that in neutral water. This gave us the impetus to investigate this pH dependent solubility further, believing it could lead to a level of Amino–G incorporation higher than that previously observed.

8.4 Future work involving GANP

Upon reviewing our studies into both the synthesis and use of GANP we feel there is little that could be added to our current results. The data collected for the levels of incorporation achieved using both crude and purified GANP are in agreement with our original hypothesis that the introduction of a phosphoamidate group would improve the incorporation with respect to Amino–G (as used by Suga *et al*).

Further transcription reaction experiments could be justified if a completely pure sample of GANP were isolated. However given the depth of our work into developing a suitable purification method, with only limited success, achieving this goal may be very difficult. The purification would be aided if the product distribution of the amino–phosphorylation step were improved to give ~100 % of the phosphoramidate, eliminating the need for strong anion exchange chromatography to remove residual starting material and by–products. The high level of conversion of Amino–G to GANP (90 %) in the presence of other nucleophiles suggests that the reaction to form the phosphoramidate is more favourable than the side–reactions to form phosphate and phosphate ester impurities. From this we can hypothesise that the reaction of $POCl_3$ with other nucleophiles, *i.e.* water, hydroxide, 2′, 3′–hydroxyls, could occur as a product of poor mixing within the reaction vessel. That is, that upon addition of $POCl_3$ to the aqueous solution the bulk of the phosphorylating agent reacts with all the Amino–G in the local area. However as the mixing is poor $POCl_3$ species are present in a high local concentration and not brought into contact with more amine on a timescale fast enough to avoid the otherwise unfavourable reactions with other nucleophiles in close proximity. By improving mixing in the vessel the distribution of $POCl_3$ throughout the reaction mixture could be made more uniform, resulting in no high local concentrations of $POCl_3$ and the product distribution of the reaction moving closer to 100 % GANP.

On a laboratory scale, improvements in the mixing energy supplied could be made by:

1. Using a baffled reaction vessel. Baffled reactors are commonly used in industrial scale processes where efficient mixing can be paramount to the success of a reaction. Baffles divert circular flow patterns, breaking up flow and reducing the mixing time of the vessel.

Given the highly reactive nature of $POCl_3$, and the high concentration of nucleophilic species in the reaction mixture, a baffled reactor may not sufficiently reduce the mixing time of the reaction. Alternative methods could be sought:

2. Using a flow processing method. As the solutions of Amino–G and $POCl_3$ are both free–flowing and miscible it would be possible to carry out the reaction using a continuous flow method whereupon the two reagents are pumped through a tubular system, meeting at a mixing device. This device could be varied as dictated by the needs of the chemistry *i.e.* the intensity of the mixing required would be determined by the $t_{1/2}$ of the reactions occurring. The reaction would be completed in a 'residence time unit' that would likely be a tube of given length or a stirred vessel. Continuous flow systems for laboratory use are commercially available.

Although using the formation of GANP as an example, the amino–phosphorylation reaction in aqueous media currently gives very good selectivity and could be readily applied to more general range of amine containing species. If such a study were undertaken improvements in selectivity and therefore ease of purification of reaction products could potentially be made using either of the methods set out above.

8.5 Amino–G Summary

As discussed above (Section 8.3) the introduction of an ionisable phosphoamidate group to Amino–G increased the nucleotide's aqueous solubility, thereby increasing the level of incorporation above that observed by Suga *et al*[45]. However, during the synthesis of GANP it became evident that the solubility of Amino–G in aqueous solution increased with pH (Section 2.50). After investigating the limits of Amino–G's solubility at various pH levels using UV spectroscopy we extended our research to include an assessment of the levels of incorporation observed when using an amine stock pre–dissolved in a $NaOH_{(aq)}$ solution (Section 6.2). As Fig:8.2 illustrates, the levels of Amino–G incorporation we observed improved upon not only the literature values, but were also higher than those recorded for GANP. We had expected that the inclusion of excess NaOH to the transcription reaction, *via* the Amino–G stock, would result in a decrease in the enzyme's activity and consequently total RNA yield of the reactions. However, we did not observe a significant reduction in RNA yield that could not be attributed to the effect of adding excess novel nucleotide to the reaction.

We undertook a range of experiments in an attempt to assess the effect that the additional hydroxide had on the pH of the transcription reaction. Unfortunately these were largely inconclusive, although we were able to make informed opinions on the pH of Amino–G containing transcription reactions and the behaviour of the Amino–G under those conditions.

- In a transcription reaction containing Amino–G some, but crucially not all, of the amine added to the reaction precipitates. The concentration remaining in solution is higher than that achieved by dissolving solid Amino–G in the reaction mixture, as indicated by our UV spectroscopy study (Section 6.2).

The pH of the reaction mixtures is buffered by a range of substances (T7 RNAP and RNasin enzymes and the TRIS of the reaction mixture) to a value close 8, as indicated by the lack of reduction in activity of the T7 RNAP enzyme. As we did not possess the time or equipment to measure the precise pH of transcription reactions that include a range of Amino–G concentrations we cannot make more definitive statements.

The route to 5′–AminoRNA using Amino–G is in many ways more successful than that developed for GANP. The first point to note is that the synthesis of Amino–G is shorter than that of GANP by one step. This is especially significant if a high level of incorporation is required, with Amino–G up to 88 % incorporation can be observed without much extra work, however to achieve anywhere near this with GANP an extensive work–up procedure must be used to acquire a purified sample. The reason behind the higher total RNA yields and levels of incorporation observed using Amino–G rather than GANP have not been thoroughly explored owing to time constraints, we can however propose some possibilities:

- Owing to the presence of a 5′–anionic group GANP is a better mimic for GTP than Amino–G, leading to an increased amount of competition during the chain elongation stage of polymerisation. As GANP does not posses the pyrophosphate

leaving group necessary to be incorporated into the 'body' of the RNA sequence this competition has the effect of reducing the total RNA yield of the reaction, relative to that of a 'standard' non–GANP containing reaction.

- As demonstrated in the works of Benkovic and Jencks, phosphoamidate groups can take part in phosphoryl transfer reactions with nucleophilic groups (Section 2.15–22). Large peptides such as the T7 RNA polymerase enzyme contain a variety of amino acid residues, some of the side–chains of which possess nucleophilic groups e.g lysine, arganine. Therefore it is possible that in the active–site of the enzyme, where GANP would have a significant residence time, a phosphoryl transfer reaction could occur between the phosphoamidate of GANP and an amino acid residue key to the enzyme's function. This phosphorylation could temporarily remove the enzyme's ability to transcribe RNA, thereby reducing the total RNA yield of the reaction.

8.6 Further work involving Amino–G

The most obvious and important area of work to be carried out regarding the use of Amino–G is the measurement of the pH levels of transcription reactions containing a variety of concentrations of the amine. To make this study complete the use of buffers of fixed pH would be advised to assess the products of transcription reactions occurring in a variety of known conditions. As we have determined the optimum amount of Amino–G to use in transcription reactions it is unlikely that the collection of this data would be used to improve either the total RNA yield or level of amine incorporation. However, it would serve to complete our knowledge of what effect Amino–G has on the pH of a transcription reaction (see Section 6.14). Other than the reagents and equipment necessary for carrying out transcription reactions we would also require a Microtip pH electrode to make the measurements, as the reaction volume required to make pH measurements using a larger electrode prohibits the inclusion of T7 RNAP and RNaisin enzymes on a cost basis. As discussed previously the enzyme species could provide significant buffering capacity, resulting in an altered pH with respect to reaction mixtures in which the enzymes are absent.

As discussed in previous sections (6.14) we expect that the true pH levels of Amino–G transcription reactions are much lower than those measured in our previous experiments (Section 6.11) in which T7 RNAP and RNasin were absent. From literature values of the pH dependent total RNA yield of T7 RNAP mediated transcription reactions we would expect that the pH measurements would reveal values in the range 8–10.

8.7 Azido–G summary

The results observed during our study of Azido–G came under two categories:

- Production of 5′–AminoRNA
- Production of 5′–AzidoRNA

Of these the most significant findings have been in the production of AzidoRNA. The methodology used to synthesise 5′–AminoRNA using Azido–G, though improving on the levels of incorporation previously reported in the literature, are not particularly relevant considering the more successful methods developed during our research.

We were only partially successful in our aim to prove the azide group could be used as a 'handle' for bioconjugation. Although we were able prove bioconjugation could be achieved our methodology was imperfect in that we could not determine the level of Azido–G incorporation owing to poor resolution of the PAGE gel used in our streptavidin assay. Owing to the absence of azido or alkyne groups in nature bioconjugation to 5′–AzidoRNA *via* a CuAAC 'click' chemistry methodology has the possibility of being very highly specific. This characteristic coupled with the ease of 5′–AzidoRNA synthesis demonstrated in our study makes the use of a direct coupling methodology potentially suitable for a number of applications.

8.8 Further work using Azido–G

To make best use of Azido–G's ability to position a reactive azide group at the 5′–terminus of RNA a method of 'click' coupling should be developed, thereby, opening a route to the highly specific bioconjugation of a wide range of groups to the RNA

terminus, limited only by the ability to attach suitable alkyne functionality to the desired group. Supporting evidence that such a methodology could be developed can be found in a recent publication of Sagheer *et al* in which a triazole containing ligand was used to tune the reactivity of Cu (I) in a CuAAC reaction so as not to degrade the DNA reagents (Fig:8.3)[67].

Fig:8.3 The self–ligation of a DNA derivative *via* a Cu (I) mediated CuAAC reaction in which the reactivity of the metal catalyst is tuned using a polyhydroxypropyl triazole amine ligand to avoid degradation of the DNA[67].

The triazole containing ligand used by Sagheer *et al* is a member of a family of ligands developed by Chan *et al* to improve selectivity in CuAAC reactions[68]. The wide range of ligands available to be used gives support to the likelihood that a methodology for the 'click' bioconjugation to 5′–AzidoRNA could be successfully developed.

9.0 Experimental Section

9.1 General Methods

Reagents

All reagents were purchased from Sigma, Aldrich, Promega or Molecular Probes and in the majority of cases used as supplied.

- Phosphorus oxychloride was distilled under vacuum and stored in a phosphorus pentoxide dessicator prior to use in reactions.
- Pyridine was distilled over calcium hydride.
- DMF and NMP were both dried over molecular sieves prior to use.

Infrared spectroscopy

Infrared spectra were recorded as solids using a Golden gate (ATR) on a Perkin–Elmer FT–IR 1600 spectrometer.

NMR spectroscopy

^{1}H, ^{13}C, and ^{31}P NMR spectra were collected on a range of instruments:

Varian Mercury 200 (^{1}H at 199.975 MHz, ^{13}C at 50.289 MHz, ^{31}P at 80.958)

Varian Unity 300 (^{1}H at 299.908 MHz, ^{13}C at 75.412 MHz, ^{31}P at 121.405 MHz)

Bruker 400 (^{1}H at 400.130, ^{13}C at 100.623)

Varian 500 (^{1}H at 499.772, ^{13}C at 125.675, ^{31}P at 202.310)

Signal assignments are reported as follows:

Chemical shift (ppm) (number of nuclei, multiplicity of signal, coupling constant *J* (Hz), assignment).

All chemical shifts are quoted relative to an internal standard of Tetramethylsilane (0 ppm) present in the deuterated solvents used in the NMR spectroscopy experiments.

Mass spectroscopy

All mass spectra were collected on a Micromass LCT spectrometer.

Liquid chromatography

Two systems were used over the course of my research the first of which (System A) consisted of:

a LKB Bromma 2120 Varioperpex® II peristaltic pump,
a Cecil Instruments CE 212 Variable Wavelength Ultraviolet Monitor (260 nm),
a LKB Bromma 2070 Ultrorac® II fraction collector,
and a Phillips PM 8262 XT recorder.

We used one of two columns with this system:

- An ion exchange column of dimensions 420 x 15 mm, packed with DEAE Sephadex® A25 media. Optimum flow–rate approximately 1.3 ml/min.

- A gel filtration column of dimensions 950 x 18 mm, packed with Sephadex® G10 media. Optimum flow–rate approximately 2.1 ml/min.

System B was an ÄKTA Prime plus liquid chromatography instrument, which housed a pump, a fraction collector and a flow UV detector (260 nm). This system was controlled *via* a PC running Primeview 5.0 and Evaluation software.

Two columns were used with this system:

- An ion exchange column packed with Capto Q strong anion exchange media (25 mL). Optimum flow–rate 5 mL/min.

- A gel filtration column packed with Sepharose G10 media (150 mL). Optimum flow–rate 5 mL/min.

Analysis of ^{32}P labelled RNA

All ^{32}P labelled RNA samples were detected through exposure to Fujifilm image plates, which were subsequently developed and analyzed using a Fujifilm FLA 300 IP reader running Aida image analyzer v.3.11 software.

9.2 Preparation of 5'-deoxy-5'-iodoguanosine[54]

Iodine (6.67 g, 26 mmol) was added over 5 minutes to a magnetically stirred suspension of guanosine hydrate (2.5 g, 8.83 mmol), triphenylphosphine (7.3 g, 27.8 mmol), imidazole (3.75 g, 55.2 mmol) and *N*-methyl pyrrolidinone (34 ml, dried over activated molecular sieves) in a round-bottomed flask (500 ml) fitted with a calcium chloride drying tube and covered with aluminium foil in order to exclude light. The pale yellow solution that formed was stirred for 3 h at room temperature. Over this period complete dissolution of the solid occurred. Dichloromethane (335 ml) and water (100 ml) were then added to the reaction vessel, this caused the precipitation of a white solid. The flask was stoppered and placed in a refrigerator for 48 h in order to maximise the amount of solid formed. Crude white solid was collected on a Büchner funnel before being refluxed in dichloromethane (100 ml) for approx. 1 h. The white product was collected in a Büchner funnel and dried at 95 °C in an oven. This afforded 5'-deoxy-5'-iodoguanosine (2.36 g, 68%) as a pale off-white powder. mp=194–198 °C (dec.); (lit.,[1] 190–200 °C [dec.]); (Found C, 30.5; H, 3.1; N, 17.8. $C_{10}H_{12}IN_5O_4$ requires C, 30.6; H 3.1; N, 17.8 %); ν_{max}(KBr disc)/cm^{-1} 1704s (C=O); ^{1}H NMR, δ_H(300 MHz, DMSO-d_6; Me_4Si): 3.55 (1 H, *ABX* system, J_{AB} 10.5, J_{AX} 6.3, 5'–CH_AH$_B$), 3.40 (1 H, *ABX* system, J_{AB} 10.5, J_{BX} 6.9, 5'-CH$_A$$H_B$), 3.86–3.96 (1 H, m, 4'-H$_X$), 4.01–4.08 (1 H, m, 3'-H), 4.61 (1 H, q, *J* 5.4, 2'-H), 5.38 (1 H, d, *J* 4.5, 3'-OH), 5.56 (1 H, d, *J* 6, 2'-OH), 5.72 (1 H, d, *J* 6, 1'-H), 6.48 (2 H, br s, NH_2), 7.92 (1 H, s, 8-H), 10.7 (1 H, br s, NH); δ_C(100.6 MHz, DMSO-d_6; Me_4Si): 8.0 (5'-C), 72.7 (3'-C), 73.1 (2'-C), 83.7 (4'-C), 86.5 (1'-C), 116.7 (5-C), 135.8 (8-C), 151.4 (4-C), 153.6 (2-C), 156.7 (7-C); *m/z* (ES$^+$) 416 ([M–Na]$^+$).

9.3 Preparation of 5'-azido-5'-deoxyguanosine[46]

Sodium azide (0.577 g, 8.88 mmol) was added to a magnetically stirred suspension of 5'-deoxy-5'-iodoguanosine (1.51 g, 3.8 mmol) and *N,N*-dimethylformamide (12 ml dried over activated molecular sieves) in a 100 ml round-bottomed flask fitted with an air condenser under a nitrogen atmosphere. The dark yellow mixture was stirred for 20 h at 80 °C in a thermostatically controlled oil bath. The reaction mixture was then allowed to cool to room temperature. The DMF was removed under reduced pressure to give a pale brown solid. Water (25 ml) was added to the residue and the mixture was stirred for 30 mins in order to remove residual sodium azide. Off-white crude product was collected on a Hirsch funnel and washed with cold water (2 × 10 ml) to remove any remaining inorganic azide. The solid was washed with cold ethanol (7 ml) to remove the water and diethyl ether (5 ml) to remove the ethanol. The white powder was dried in a vacuum desiccator to give 5'-azido-5'-deoxyguanosine (0.915 g, 78%): mp=198–205 °C (dec.); (Found C, 38.9; H, 3.9; N 36.35. $C_{10}H_{12}N_8O_4$ requires C, 39.0; H, 3.9; N, 36.35%); ν_{max}(KBr disc)/ cm^{-1} 2108s (N_3); ^1H NMR, δ_H(300 MHz, DMSO-d_6; Me_4Si): 3.52 (1 H, *ABX* system, J_{AB} 13.2, J_{AX} 7.2, 5'–CH_AH$_B$), 3.66 (1 H, *ABX* system, J_{AB} 13.2, J_{BX} 3.9, 5'-CH$_A$$H_B$), 3.94–4.02 (1 H, m, 4'-H$_X$), 4.01 (1 H, t, *J* 4.4, 3'-H), 4.58 (1 H, t, *J* 5.4, 2'-H), 5.73 (1 H, d, *J* 6, 1'-H), 6.53 (2 H, br s, NH_2), 7.90 (1 H, s, 8-H): δ_C(100.6 MHz, DMSO-d_6; Me_4Si): 51.8 (*5'-C*), 70.9 (*3'-C*), 72.6 (*2'-C*), 82.8 (*4'-C*), 86.8 (*1'-C*), 117.2 (*5-C*), 135.8 (*8-C*), 151.8 (*4-C*), 153.9 (*2-C*), 157.0 (*7-C*); *m/z* (ES^+) 331.1 ($[M + Na]^+$).

9.4 Preparation of 5'-amino-5'-deoxyguanosine[46]

Triphenylphosphine (1.69 g, 6.45 mmol) was added to a magnetically stirred suspension of 5'-azido-5'-deoxyguanosine (1.00 g, 3.24 mmol) and dry pyridine (16.2 ml) in a 250 ml round-bottomed flask at 0 °C with a $CaCl_2$ drying tube fitted. After stirring at room temperature for 3 h the suspension thickened considerably. The mixture was re-cooled to 0°C and cold ammonium hydroxide solution (5 ml; 0.880 ammonia : 17 ml water) was added. The mixture was then stirred for a further 18 h. Solvents were removed under

reduced pressure. The remaining pale brown crude product was suspended in ethyl acetate (95 ml) and stirred at room temperature for 15 mins. Solids were collected at the pump and then washed with cold ethyl acetate (18 ml), cold ethyl acetate: methanol (50:50, 18 ml) before recrystallising from water to give off-white crystals of 5'-amino-5'-deoxyguanosine (0.557 g, 74%). mp=218–220 °C (dec.); lit.,[2] 219–220 °C; (Found C, 39.8; H, 5.4; N, 27.7. $C_{10}H_{14}N_6O_4.H_2O$ requires C, 40.0; H, 5.3; N, 28.0%); ν_{max}(KBr disc)/cm^{-1} 2622br (NH_2), 1702s (C=O); ^{1}H NMR, δ_H(500 MHz, DMSO-d_6; Me_4Si): 2.72 (1 H, *ABX* system, J_{AB} 13.5, J_{BX} 5.5, 5'–CH_AH$_B$), 2.77 (1 H, *ABX* system, J_{AB} 13.5, J_{AX} 4.5, 5'-CH$_A$$H_B$), 3.77–3.81 (1H, m, 4'-H$_X$), 4.08 (1H, t, *J* 4.5, 3'-H), 4.44 (1H, t, *J* 6.8, 2'-H), 5.66 (2H, d, *J* 6.5, 1'-H), 6.52 (1H, br s, NH_2), 7.93 (1H, s, 8-H): δ_C(100.6 MHz, DMSO-d_6; Me_4Si): 43.5 (*5'-C*), 70.6 (*3'-C*), 73.11 (*2'-C*), 85.5 (*4'-C*), 86.2 (*1'-C*), 116.7 (*5-C*), 135.71 (*8-C*), 151.3 (*4-C*), 153.6 (*2-C*), 156.7 (*6-C*)): *m/z* (ES$^+$) 283.1 ([M + H]$^+$).

9.5 Modified Hampton protocol for GANP synthesis[47]

5'-amino-5'-deoxyguanosine (300 mg, 800 nmol) was added to a magnetically stirred mixture of acetonitrile (1 mL, 19 mmol, dried over molecular sieves), pyridine (100 μL, 1.2 mmol, distilled off Calcium hydride) and phosphorus oxychloride (420 μL, 1.35 mmol) at 0–5 °C. The resulting suspension was stirred for 48 h at a maintained 0–5 °C, forming a thick brown paste. The reaction mixture was added to a saturated solution of Barium hydroxide in water then stirred for 0.5 h. The suspension was then centrifuged to separate the solid barium phosphate from the barium salt of GANP in solution. Addition of ethanol (2 volumes) to the solution resulted in the precipitation of a white solid that was separated from the solution by further centrifugation. The solid was then dried under vacuum before attempts were made to dissolve it in D_2O, however only a small amount of the solid was soluble. Only ^{1}H NMR showed signals, however they were of too low intensity to assign.

9.6 Modified Hampton protocol for GANP synthesis (2)[47]

Cold phosphorus oxychloride (15 μL, 1.61 mmol) was added to a magnetically stirred suspension of 5'-amino-5'-deoxyguanosine (50 mg, 177 nmol) in acetonitrile (40 μL,

7.62 mmol, dried over mol. sieves) and pyridine (4 μL, 49 nmol) at 0–5 °C. The mixture was then left to stir for 72 h, after this period the reaction was quenched through addition of a sodium carbonate solution (60 mL, 0.5 M), forming a yellow solution of pH 8. The solution was then loaded on to a liquid chromatography column packed with A25 DEAE Sephadex ion exchange media (System A,). A Triethylammonium bicarbonate (TEAB) eluent (0.1 – 0.5 M, pH 7.5) was used to separate compounds in the reaction mixture. Those fractions containing UV active compounds were lyophilised to give white solids. ^{1}H NMR and mass spectroscopy of the samples shows evidence of high levels of triethylamine present, but no other organic compounds.

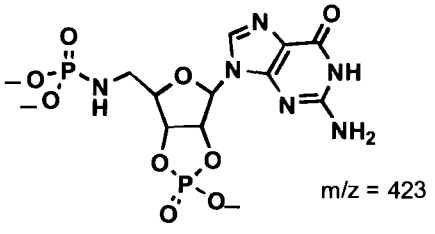
9.7 Modified Hampton protocol for GANP synthesis (3)[47]

Cold phosphorus oxychloride (15 μL, 1.61 mmol) was added to a magnetically stirred suspension of 5′-amino-5′-deoxyguanosine (50 mg, 177 nmol) in acetonitrile (40 μL, 7.62 mmol, dried over mol. sieves) and pyridine (4 μL, 49 nmol) at 0–5 °C. The mixture was then left to stir for 72 h, after this period the reaction was quenched through addition of a sodium carbonate solution (60 mL, 0.5 M), forming a yellow solution of pH 8. The solution was then loaded on to a liquid chromatography column packed with A25 DEAE Sephadex ion exchange media (System A). An ammonium bicarbonate eluent (0.1 – 0.5 M, pH 7.5) was used to separate compounds in the reaction mixture. Those fractions containing UV active compounds were lyophilised to give white solids. The solids were redissolved in a small amount of water before loading on to a liquid chromatography column packed with Sephadex G10 size exclusion media, separation of products was carried out using a water eluent (System A). The resulting UV active fractions were lyophilised to give white solids. ^{1}H NMR of the samples featured large water peaks with a number of inconclusive low intensity signals.

9.8 Modified Yoshikawa protocol for GANP synthesis[55,56]

A mixture of phosphorus oxychloride (135 μL, 0.53 mmol), trimethylphosphate (1.31 mL, 6.7 mmol) and pyridine (15 μL, 200 nmol, distilled off Calcium hydride) was cooled

to 0 °C before being added to 5′-amino-5′-deoxyguanosine (55 mg, 172 nmol) in a stoppered round bottomed flask. The reaction mixture was stirred for 16 h before being added to a solution of sodium perchlorate (215 mg, 1.188 mmol) in dry acetone (21.3 mL) and ether (10.7 mL) resulting in the formation of a white precipitate. The solution was then left to stir for a further 8 h, before the solid was separated by centrifugation and ddecanting of the supernatant. The solid was then washed with acetone (2 x 35 mL) and ether (35 mL) and dried under vacuum over phosphorus pentoxide to give a white powder (69.4 mg). Ion exchange chromatography (A25 DEAE Sephadex, TEAB buffer, 0.1–0.5 M, pH 7.5, System A) was used to purify the crude product. The UV active fractions were lyophilised to give white solids. Solid collected from one UV peak in the chromatogram was found to include phosphorus–containing species other than trimethylphosphate. ^{31}P NMR, δ_P(121 MHz, H_2O): 21.0 (s, $OPO(OR)_2$), 9.73 (s, P–N) 3.81 (s, $OP(OMe)_3$): m/z (LC–ES$^-$) 423 ([(**1**)+H]) 343 ([(**1**)–$\mathbf{OPO_3}$])[54].

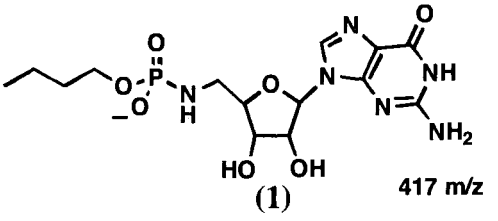


(**1**)

9.9 Yoshikawa protocol incorporating n–Butanol[55]

Phosphorus oxychloride (136 μL, 530 nmol), trimethylphosphate (1.306 mL, 6.7 mmol) , triethylamine (30 μL, 400 nmol) and 1-butanol (21 μL, 220 nmol) were mixed in a round bottomed flask and cooled to 0 °C. This solution was then added to 5′-amino-5′-deoxyguanosine (50 mg, 172 nmol) in a separate flask to form an off-white suspension. The reaction was stirred for 2 h at 0 °C, during which time dissolution of the solid occurred, forming a pale yellow solution. The reaction was quenched by addition to a solution of triethylamine (1 mL) in water (17 mL). The resulting solution had the majority of the trimethylphosphate removed through washes with chloroform (2 × 30 mL). The aqueous phase was diluted with water (to 30 mL) and had its pH adjusted to

7.5 before being loaded on to an A25 DEAE Sephadex ion exchange column (TEAB gradient 0.1-0.5 M, pH 7.5, System A). The major UV active fractions were collected, pooled and lyophilised. Solid collected from a single UV trace in the chromatogram was found to incorrporate phosphorus containing material other than trimethylphosphate. ^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 10.06 (s, P–N), 3.42 (s (br) $OP(OMe)_3$): m/z (LC–ES^-)417 ([**(1)**+H]).



(1) 417 m/z

9.10 First attempt at GANP synthesis *via* an aqueous method[58]

5′-amino-5′-deoxyguanosine (20 mg, 68.8 μmol) was dissolved in DMSO (500 μL, 5.82 mmol), the clear solution formed was then diluted with water (500 μL) and cooled to 0 °C. A dropwise co-addition was then made of phosphorus oxychloride (7.04 μL, 75 μmol) diluted in dry THF (34.4 μL) and sodium hydroxide (109 mg, 2.7 mmol) in water (1.1 mL). As these additions were made the pH was repeatedly checked to ensure the solution remained above pH 10. When full addition of the phosphorus oxychloride solution had been made the reaction mixture was diluted with water (to 750 μL) to give a clear colourless solution. ^{31}P NMR and mass spectra collected were collected. ^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 8.52 (t, $OP(O)_2$–NH–CH_2), 2.31 (s, OPO_3): m/z (LC–ES^-) 361 ([**GANP**+H]).

9.11 GANP synthesis – DMSO study

Six phosphorylation reactions were carried out containing the identical amounts of 5′-amino-5′-deoxyguanosine,sodium hydroxide, and using the same quantity of phosphorus oxychloride. The solvent proportions used were varied in each reaction as shown in the table below.

% DMSO (v/v)	DMSO (μL)	H_2O (μL)
0	0	400
10	40	360
20	80	320
30	120	280
40	160	240
50	200	200

The general procedure for the reactions was as follows:

5′-amino-5′-deoxyguanosine (10 mg, 35 μmol) was added to DMSO (see table above for quantites) and stirred for 5 mins, over which time full dissolution of the solid occurred. Water was added to the solution, in most cases resulting in the precipitation of a white solid. The reaction vessel was then cooled to 0 °C (in an ice–bath) before addition of $NaOH_{(aq)}$ (3.75 mmol) resulted in the dissolution of all solid to give a clear solution. Additions of phosphorus oxychloride solution (10 μL [22.8 μL $POCl_3$ in 77.2 μL dry THF]) were made at 60 sec intervals for 10 minutes. After all additions had been made the solvents were removed under reduced pressure to give a white solid.

All crude products were redissolved in H_2O (1 mL) and used to collect ^{31}P NMR spectra.

DMSO (%)	GANP (%)	Phosphate (%)
50	48.23	51.77
40	55.02	44.98
30	61.74	38.26
20	48.14	51.86
10	60.9	39.1
0	69.23	30.77

Typical ^{31}P NMR assignments.

^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 8.5 (t, $OP(O)_2$–NH–CH_2), 2.67 (s, OPO_3)

9.12 GANP synthesis – hydroxide study

5′-Amino-5′-deoxyguanosine (10 mg, 35 μmol) was added to DMSO (see table above for quantites) and stirred for 5 mins, over which time full dissolution of the solid occurred. Water was added to the solution, in most cases resulting in the precipitation of a white

solid. The reaction vessel was then cooled to 0 °C (in an ice–bath) before addition of $NaOH_{(aq)}$ (3.75 mmol) resulted in the dissolution of all solid to give a clear solution. Additions of phosphorus oxychloride solution (10 μL [22.8 μL $POCl_3$ in 77.2 μL dry THF]) were made at 60 sec intervals for 10 minutes. After all additions had been made the solvents were removed under reduced pressure to give a white solid.

Six phosphorylation reactions were carried out following the procedure above, the molarity of hydroxide used was varied in each reaction as shown in the table below.

Equivalents of NaOH	moles (μmol)	Concentration of NaOH stock (M)
4	140	7.0
4.5	158	7.9
5	175	8.8
5.5	193	9.6
6	210	10.5
6.5	228	11.4

5′-Amino-5′-deoxyguanosine (10 mg, 35 μmol) was added to H_2O (200 μL) and stirred for 5 mins, forming a suspension. The reaction vessel was then cooled to 0 °C (in an ice–bath) before addition of $NaOH_{(aq)}$ (200 μL, Concentration varied as shown in table above) resulted in the dissolution of all solid to give a clear solution. Additions of phosphorus oxychloride solution (10 μL [22.8 μL $POCl_3$ in 77.2 μL dry THF]) were made at 60 sec intervals for 10 minutes. After all additions had been made the solvents were removed under reduced pressure to give a white solid.

Eqs NaOH	% GANP	% Phosphate	% Impurities
4	47	41	12
4.5	88	9	3
5	91	9	Trace
5.5	85	9	6
6	80	11	9
6.5	55	25	20

Typical ^{31}P NMR assingments.

^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 9.50 (t, O*P*(O)$_2$–NH–C*H*$_2$), 4.85 (s, OPO_3), 4.59, 3.98 (d, 2′/3′–phosphoester by–products).

9.13 Synthesis of impure GANP (Optimised NaOH eq)

5′-Amino-5′-deoxyguanosine (500 mg, 1.77 mmol) was dissolved in a mixture of $NaOH_{(aq)}$ (8.85 mL of 1 M, 8.85 mmol (5 eq)) and water (1.15 mL) then cooled to 0 °C in an ice bath. Phosphorus oxychloride (165 μL, 1.77 mmol) in anhydrous THF (5 mL) was then added dropwise to the solution over 10 mins. After addition was complete the solvents were removed under reduced pressure to give crude product as a white solid (746 mg). ^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 7.61 (t, O*P*(O)$_2$–NH–C*H*$_2$), 3.82 (s, OPO_3), 2.60, 2.20 (d, 2′/3′–phosphoester by–products). : m/z (LC–ES$^-$) 361 ([**GANP**+H]).

9.14 GANP synthesis – attempted purification by Gel filtration chromatography

5′-Amino-5′-deoxyguanosine (500 mg, 1.77 mmol) was dissolved in a mixture of $NaOH_{(aq)}$ (8.85 mL of 1 M, 8.85 mmol (5 eq)) and water (1.15 mL) then cooled to 0 °C in an ice bath. Phosphorus oxychloride (165 μL, 1.77 mmol) in anhydrous THF (5 mL) was then added dropwise to the solution over 10 mins. After addition was complete the solvents were removed under reduced pressure to give crude product as a white solid. The solid was dissolved into H_2O (2.5 mL) and loaded onto a LC column packed with gel filtration media. Material was eluted from the column using an aqueous solution of potassium carbonate (2 mM), with the elution of UV active compounds being tracked by a flow- detector. An automated fraction collector was used to separate the eluted materials into a range of samples. The chromatogram showed two peaks in the UV signal,

the fractions relating to which were pooled and lyophilised to give white solids. The collected material was redissolved in a small volume of deuterium oxide and the resulting solutions used to collect ^{31}P NMR spectra. UV peak 1 ^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 7.61 (t, $OP(O)_2$–NH–CH_2), 3.82 (s, OPO_3), UV peak 2 ^{31}P [^{1}H] NMR, δ_P(121 MHz, H_2O): 3.82 (s, OPO_3).

9.15 GANP synthesis – attempted purification through selective precipitation

5′-Amino-5′-deoxyguanosine (50 mg, 177 μmol) was dissolved in a mixture of $NaOH_{(aq)}$ (885 μL of 1 M, 885 μmol (5 eq)) and water (1 mL) then cooled to 0 °C in an ice bath. Phosphorus oxychloride (33 µL, 177 μmol) in anhydrous THF (1 mL) was then added dropwise to the solution over 10 mins. After addition was complete the solvents were removed under reduced pressure to give crude product as a white solid. The solid was redissolved in a small amount of water (750 μL) and a ^{31}P NMR collected.

∂	Integral	assignments
9.30 (t)	80	GANP
5.19 (s)	11	phosphate
4.36 (d)	4.7	phosphoester
3.79 (d)	4.7	phosphoester

Estimated conversion of $POCl_3$ as 80 % *i.e.* 142 μmol, therefore we can estimate that the sample contains 35 μmol of phosphorus containing impurities.

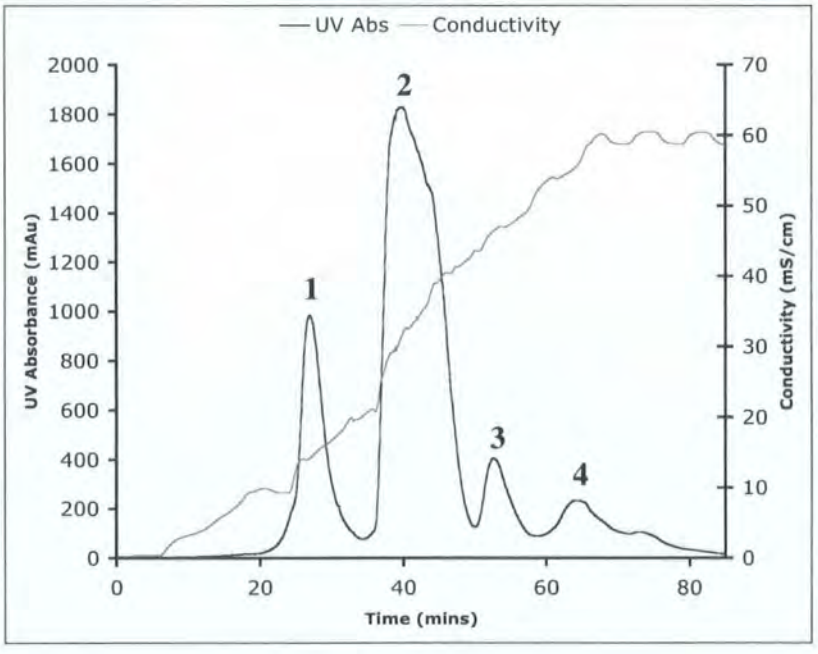
The solution of crude product had barium chloride (35 μmol) added to it, causing the precipitation of a white solid. The supernatant was separated by centrifugation and decanting. Subsequent ^{31}P NMR showed a change indicating which materials had been removed by precipitation.

∂	integral	assignment
10.99	91	GANP
6.1	5	Phosphate
5.5	2.5	phosphoester
4.8	2.4	phosphoester

Repeated precipitations and centrifugation reduces the amount of material in the sample. But does not result in a purer product.

9.16 GANP synthesis – purification by anion exchange and gel filtration chromatography

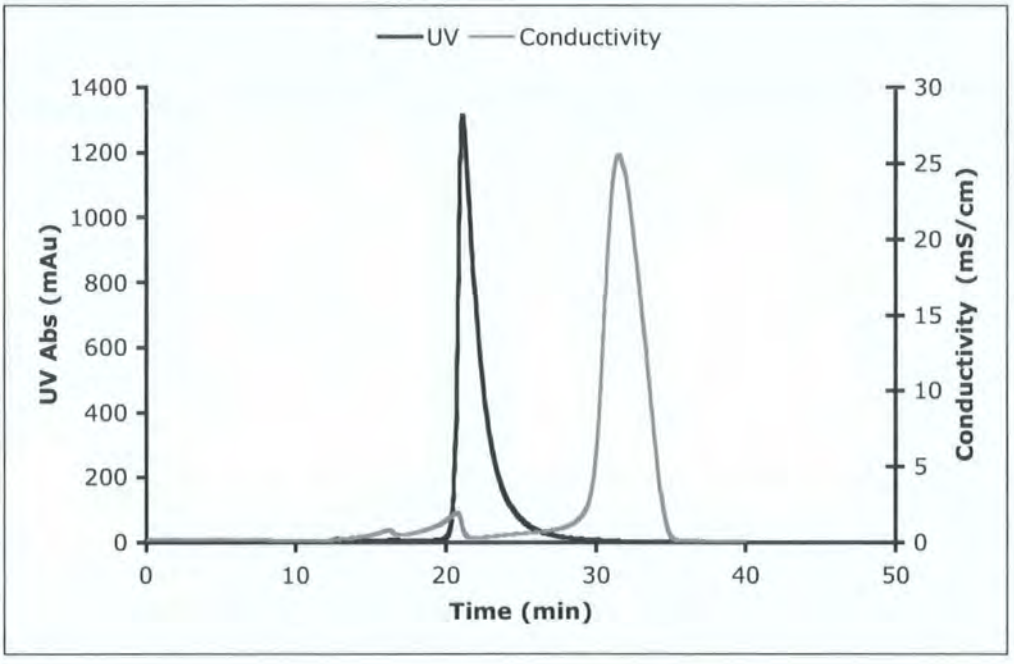
5′-Amino-5′-deoxyguanosine (50 mg, 177 μmol) was dissolved in a mixture of $NaOH_{(aq)}$ (885 μL of 1 M, 885 μmol (5 eq)) and water (1 mL) then cooled to 0 °C in an ice bath. Phosphorus oxychloride (33 µL, 177 μmol) in anhydrous THF (1 mL) was then added dropwise to the solution over 10 mins. The reaction mixture was diluted with water (3 mL) and loaded on to a Capto Q strong anion exchange chromatography column (System B). Materials were eluted from the column using a gradient of NaOH (0.1–1 M), the UV absorbance and conductivity of the eluting solution were recorded, giving the chromatogram shown below.



Fractions relating to each UV peak were collected, pooled and lyophilized to give relatively large quantities of white solids, which were subsequently redissolved in a small volume of water (750 μL) and a ^{31}P NMR spectrum was obtained.

Peak	^{31}P NMR signal	^{31}P containing species present
Crude	10.1 (t), 5.9(s)	GANP, phosphate
1	6.6(s)	Phosphate
2	9.9(t)	GANP
3	7.6(s), 5.2(s), 2.79(s)	Phosphoesters
4	None	None

The sample relating to peak 2 was then loaded on to a gel filtration chromatography column with the aim of desalting the GANP (System B).



Fractions relating to the UV signal were collected, pooled and lyophilised to give GANP as an off-white solid (54 mg, 150 μmol) characterised by mass spectrometry ^{1}H, ^{13}C, and ^{31}P NMR. ^{1}H NMR, δ_H(500 MHz, D_2O; Me_4Si): 2.86 (1 H, *ABX* system, J_{AB} 13.5, J_{AXP} 6.5, 5'–CH_AH$_B$), 2.93 (1 H, *ABX* system, J_{AB} 13.5, J_{BXP} 5.5, 5'-CH$_A$$H_B$), 4.04–4.02 (1H, m, 4'-H$_X$), 4.20 (1H, t, *J* 4, 3'-H), 4.61 (1H, t, *J* 5.5, 2'-H), 5.66 (2H, d, *J* 6.5, 1'-H), 7.74 (1H, s, 8-H): δ_C(100.6 MHz, D_2O; Me_4Si): 44.4 (*5'-C*), 71.5 (*3'-C*), 73.1 (*2'-C*), 85.4 (d, J_{CP} 10.1, 4'-*C*), 86.9 (*1'-C*), 117.9 (*5-C*), 136.2 (*8-C*), 151.8 (*4-C*), 161.1 (*2-C*), 167.8 (*6-C*): HR–MS *m/z* (ES$^-$) 361 ([M + H]$^+$).

9.17 Observing GANP hydrolysis by ^{31}P NMR spectroscopy

As the rate of GANP hydrolysis varies with pH we required a number of methods for both incubating the phosphoramidate samples at 37 °C and the required pH, and collecting ^{31}P NMR spectra at chosen time points. In all of the methods detailed below the concentrations given for buffer species are those of the total buffer species present in the reaction *i.e.* both the charged and uncharged forms. All buffers used were made up as stock solutions with the buffer species first being dissolved in a small amount of H_2O, then the pH adjusted through addition of acid or base to the desired level. After checking the pH level the buffer solution was made up to the correct dilution, before the pH was measured again.

pH 3–3.5

A solution of GANP (100 mL, 30 mM) containing sodium formate buffer (0.5 M, pH 3/3.5) was incubated at 37 °C. Aliquots (1 mL) were removed at 5 minute intervals for 30 minutes and quenched in sodium hydroxide (1 mL, 0.1 M). Each of these samples were used to collect a ^{31}P NMR spectra, the peaks of which were integrated and the collected data were tabulated and plotted.

pH 7.2–9

An NMR spectroscopy tube containing a solution of GANP (750 μL, final concentration 30 mM) and the required buffer (0.5 M):

pH	Buffer
7.2	MES
8	Bicarbonate
9	Borate

was incubated at 37 °C in a spectrometer for 12 h. During the incubation period acquisitions were made at regular intervals producing a range of ^{31}P NMR spectra, the peaks of which were integrated and the collected data were tabulated and plotted.

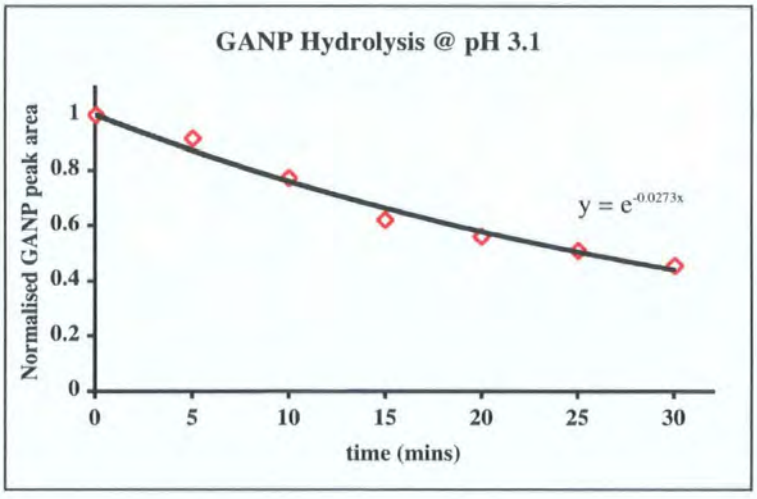
pH 9.8–10.5

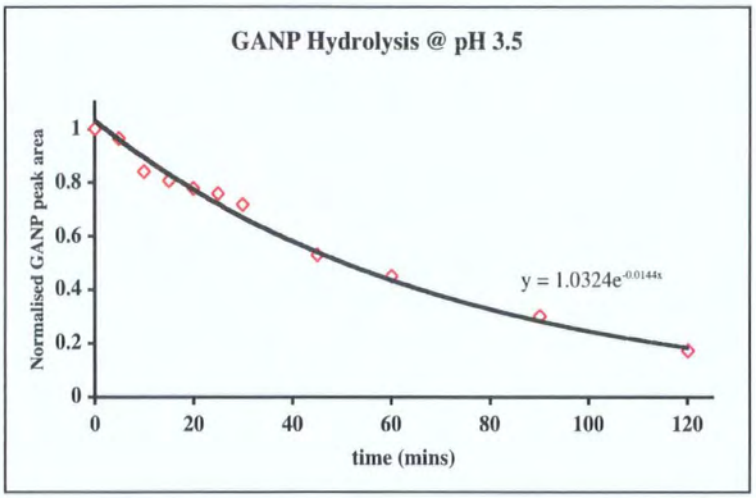
An NMR spectroscopy tube containing a solution of GANP (750 μL, 30 mM) and the required buffer (0.5 M):

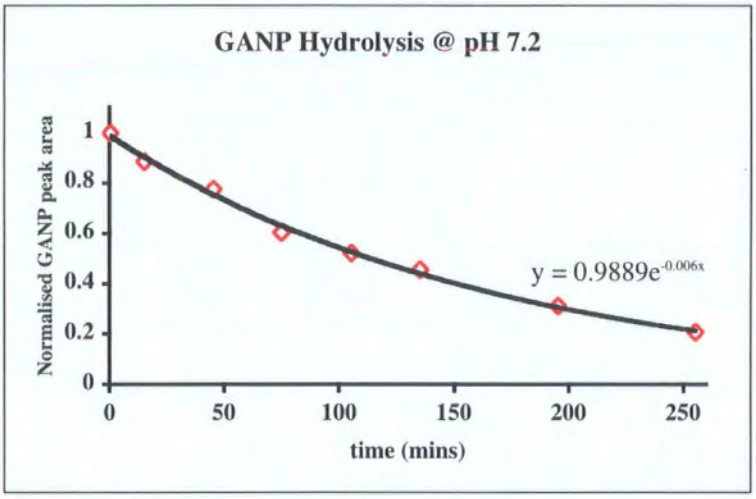
pH	Buffer/Base
9.8	Carbonate
10.5	NaOH

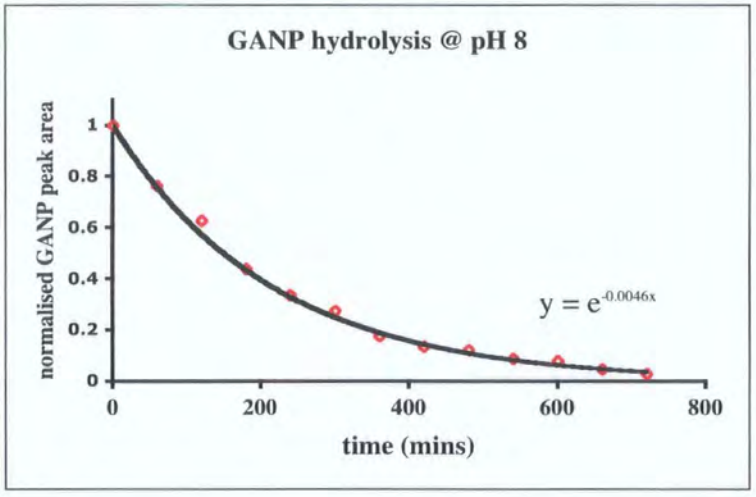
was incubated at 37 °C in a water bath for ten days. During the incubation period acquisitions were made at regular intervals producing a range of ^{31}P NMR spectra, the peaks of which were integrated and the collected data were tabulated and plotted.

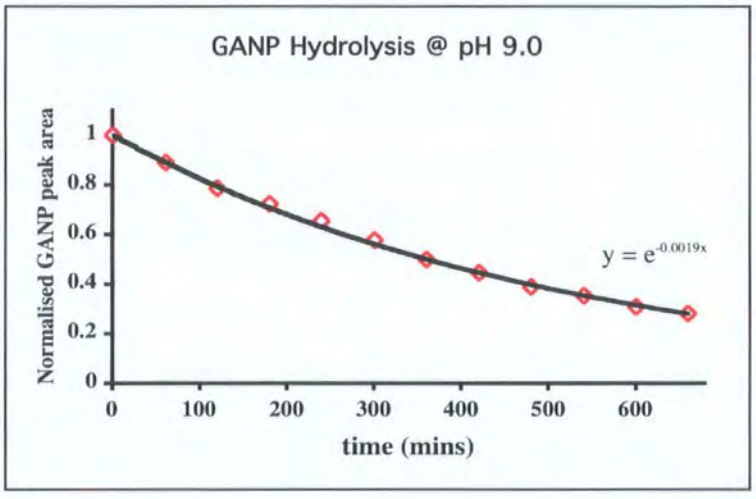
9.18 Hydrolysis data

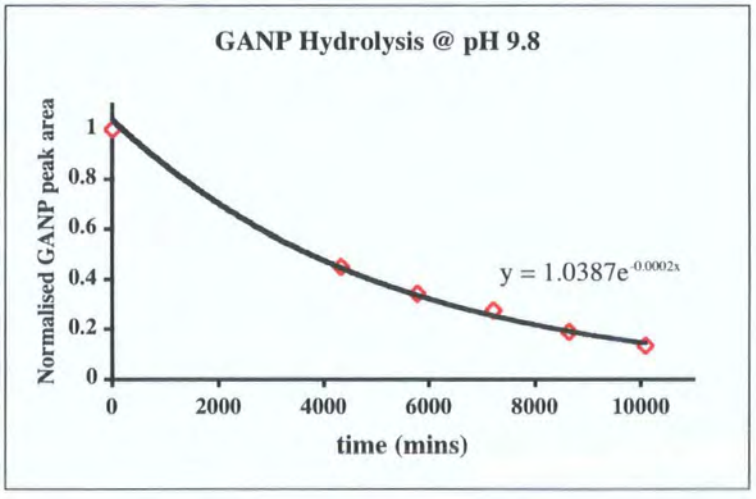


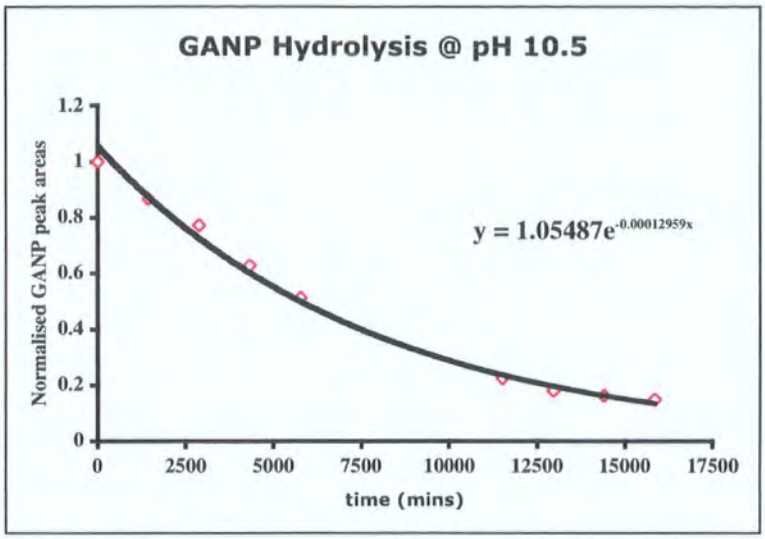

GANP Hydrolysis @ pH 3.5
Normalised GANP peak area
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
time (mins)
y = 1.0324e-0.0144x

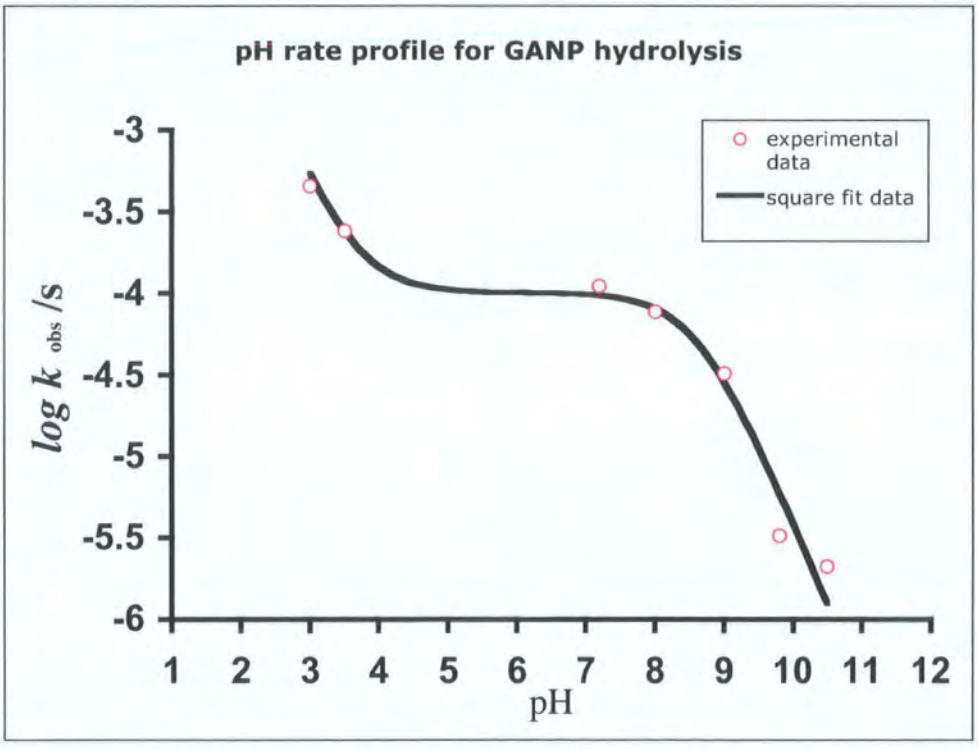

GANP Hydrolysis @ pH 7.2
Normalised GANP peak area
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
time (mins)
y = 0.9889e-0.006x


GANP hydrolysis @ pH 8
normalised GANP peak area
1
0.8
0.6
0.4
0.2
0
0
200
400
600
800
time (mins)
y = e-0.0046x


GANP Hydrolysis @ pH 9.0
Normalised GANP peak area
1
0.8
0.6
0.4
0.2
0
0
100
200
300
400
500
600
time (mins)
y = e-0.0019x


GANP Hydrolysis @ pH 9.8
Normalised GANP peak area
1
0.8
0.6
0.4
0.2
0
0
2000
4000
6000
8000
10000
time (mins)
y = 1.0387e-0.0002x





Where the equation for the square fit data is:

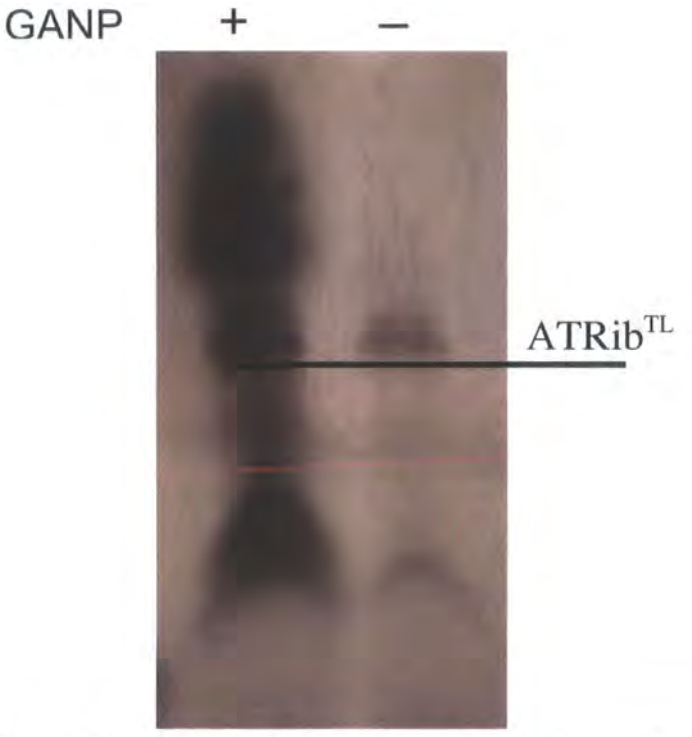
$k_{obs} = 26.7266 \times 10^{pH} + (0.00603573/ (1+10^{pH-8.599}))$

pH	k_{obs} (s^{-1})	Log k_{obs} (s^{-1})	$t_{1/2}$ (h)
3	4.55E-04	-3.3	0.42
3.5	2.40E-04	-3.6	0.80
7.2	1.11E-04	-4.0	1.7
8	7.71E-05	-4.1	2.5
9	3.21E-05	-4.5	6
9.8	3.25E-06	-5.5	59
10.5	2.10E-06	-5.7	92

9.19 First transcription using Crude GANP[3]

Two transcription reactions were carried out, both of 50 μL total volume; reaction A contained crude GANP (8 mM) while reaction B contained none. Both reactions were carried out using the methodology below.

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$)[61]. GANP (4 μL of a 100mM stock) was only added to reaction mixture A. DEPC water was added to dilute the reaction mixtures to the desired volume (50 μL). Both reaction mixtures were then incubated at 37 °C for 2.5 h. Following which the reaction was halted *via* digestion of the DNA template through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. The reaction mixtures were then diluted in loading buffer (1 volume, 8 M urea, 20 mM EDTA, 2 mM TRIS in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), and the gel visualised by UV–shadowing.



Digital photograph of UV shadowing analysis, showing species separation by PAGE.

***<u>DNA template</u>**

The dsDNA template containing a T7 promotor sequence was constructed *via* PCR using synthetically derived ssDNA template and primers. The sequences are based on those used by Suga *et al.*

Primer 1
5'–GGT AAC ACG CAT ATG TAA TAC GAC TCA CTA TAG GAA CAA CTT GCA GCT CAT T–3'

Primer 2
5'–CAA CCA AAA ACA AAA AGC AT–3'

Template ssDNA
5'– CAA CCA AAA ACA AAA AGC ATC ACG TAT GAT GCT TCT AAC CAT TTT CAC GGT AAG ATG CAG CTC CAC GAA TGG AGC TGC AAG TTG TTC C–3'

For ease of undertaking PCR reactions we used an 11X buffer stock solution, consisting of:

2 M TrisHCl (pH 8.0) = 167 μL
1 M Ammonium Sulfate = 83 μL
1 M $MgCl_2$ = 33.5 μL
2–Mercaptoethanol = 3.6 μL
10 mM EDTA (pH 8.0) = 3.4 μL
dNTP 100 mM each of dGTP, dATP, dCTP, dTTP = 75 μL
10 mg/mL Bovine serum albumin = 85 μL
Total volume = 676 μL

The 11X buffer was used in PCR reactions according to the following receipe:

11X buffer stock = 18 μL
2 μM $Atrib^{TL}$ ssDNA template = 2 μL
100 μM Primer 1 = 20 μL
100 μM Primer 2 = 20 μL
DEPC water = 136 μL
50X Taq Polymerase = 4 μL
Total volume = 200 μL

The PCR cycle followed was:

94 °C 1 min
55 °C 1 min
72 °C 1 min

On average 15 –20 cycles were required to produce enough DNA to be visualised *via* an Ethidium bromide stain in an agaraose electrophoresis gel.

9.20 Crude GANP transcription (2)[3]

A series of transcription reactions (75 µL each) was undertaken, each containing a different concentration of GANP. Our aim was to investigate whether the intensity of the diffuse signal observed in the +GANP lane of the UV–shadowing image in experiment 9.18 increased with GANP concentration.

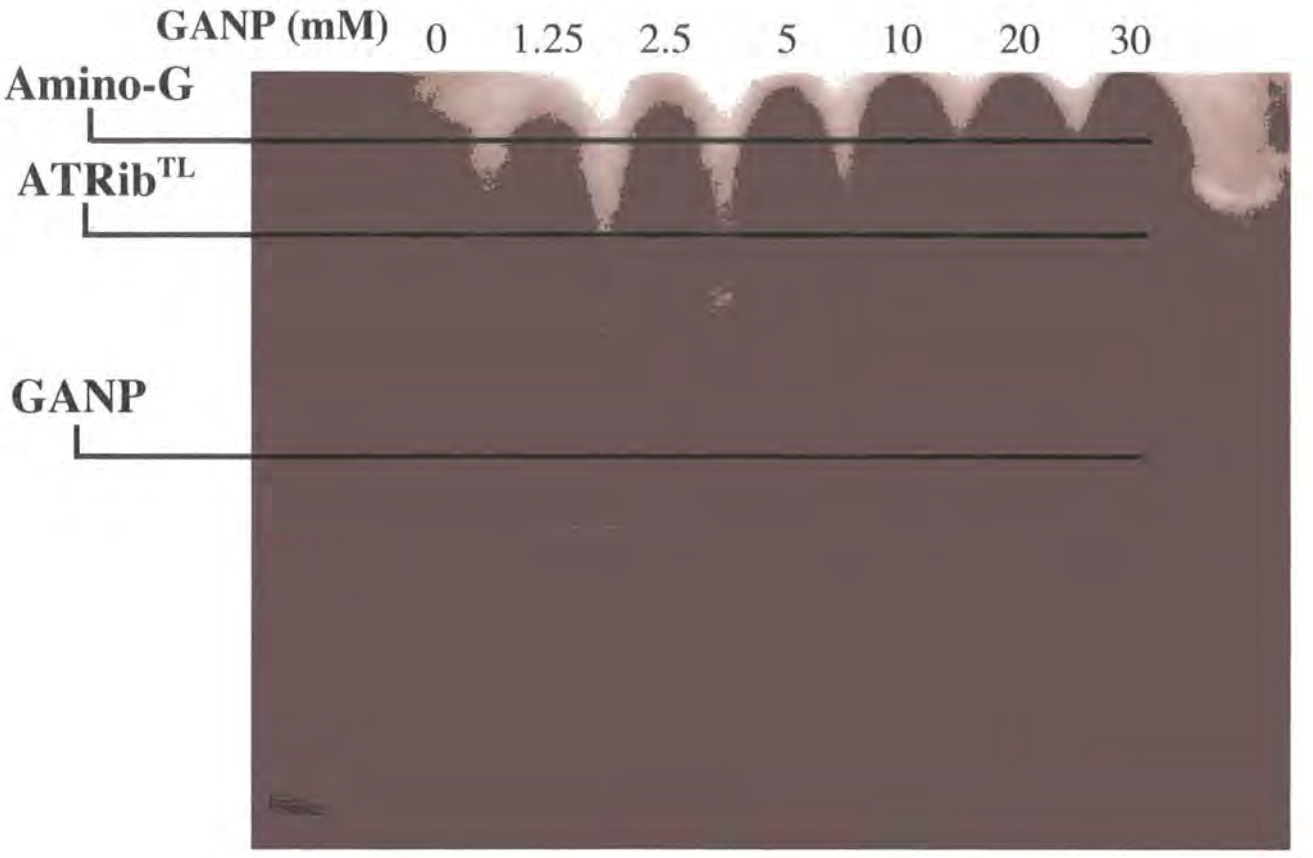
Each of the seven reaction mixtures contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of

each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5$ M^{-1} cm^{-1})[61].

The amount of crude GANP and DEPC water added to each reaction mixture was varied for each reaction according to the table below:

Reaction	100 mM crude GANP (μL)	DEPC water (μL)	Final crude GANP conc (mM)
1	0	23	0
2	0.94	22.06	1.25
3	1.88	21.12	2.5
4	3.75	19.25	5
5	7.5	15.5	10
6	15	8	20
7	23	0	30

All reaction mixtures were incubated at 37 °C for 2.5 h, following which the reaction was halted *via* digestion of the DNA template through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. The reaction mixtures were then diluted in loading buffer (1 volume, 8 M urea, 20 mM EDTA, 2 mM TRIS in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), and the gel visualised by UV–shadowing.



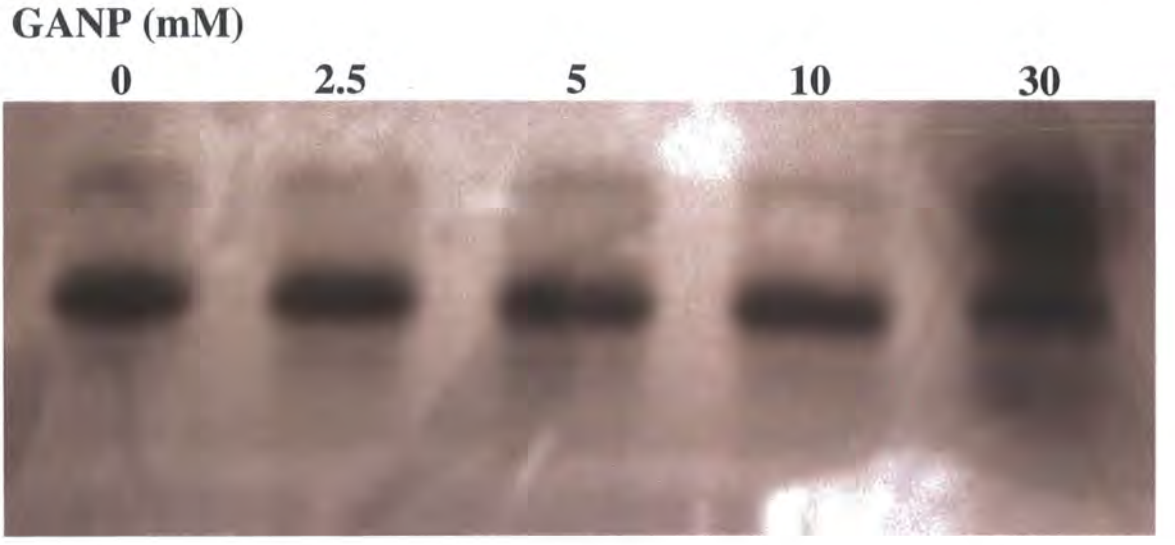
9.21 Purification of transcripts by gel filtration[3]

Five transcription reactions (75 μL) containing 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$) were made up[61].

The amount of crude GANP and DEPC water added to each reaction mixture was varied for each reaction according to the table below:

Reaction	100 mM crude GANP (üL)	DEPC water (uL)	Final GANP conc (mM)
1	0	23	0
2	0.94	22.06	1.25
3	1.88	21.12	2.5
4	3.75	19.25	5
5	23	0	30

Reaction mixtures were incubated at 37 °C for 2.5 h. The DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. Unincoporated GANP and other small molecules were removed *via* gel filtration using a spin-column (ZEBA Desalt 0.5 ml, Pierce). The material eluted from the column was then mixed with loading buffer (1 volume, 8 M urea in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (70 % ethanol) then redissolving in DEPC water.



Digital photograph of UV shadowing analysis, showing species separation by PAGE.

9.22 Transcription Reactions incorporating crude GANP[3]

On the basis the findings made in Expt. 9.20 (see Section:5.7) all subsequent transcription reactions were carried out using the general protocol set out below. The only variable between reactions was the amount of GANP used in each experiment. Reagents other than GANP and T7 RNAP were combined as a stock solution before distribution between each reaction vessel in use. In this way the amount of NTPs, [α-^{32}P]-UTP and other reagents in each reaction could be standardised across the experimental range.

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$)[61]. Crude GANP was added to the reaction mixture as a 100 mM stock and diluted appropriately with DEPC water to give the desired final concentration. Reaction mixtures were incubated at 37 °C for 2.5 h. The DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. Residual GANP and other small molecules were removed *via* gel filtration using a spin-column (ZEBA Desalt 0.5 ml, Pierce). The material eluted from the column was then mixed with loading buffer (1 volume, 8 M urea in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea)

polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (7:3 ethanol:water) then redissolving in DEPC water.

Transcription reactions performed on a 75 μL scale in the absence of novel nucleotide yielded 300-350 pmol of RNA.

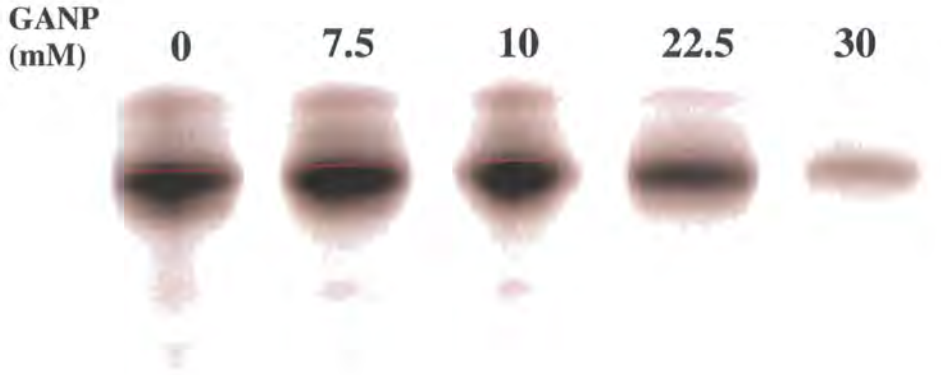
9.23 Amine–specific biotinylation reactions

In all the studies carried out the amine–specific biotinylation reactions followed the general protocol detailed below.

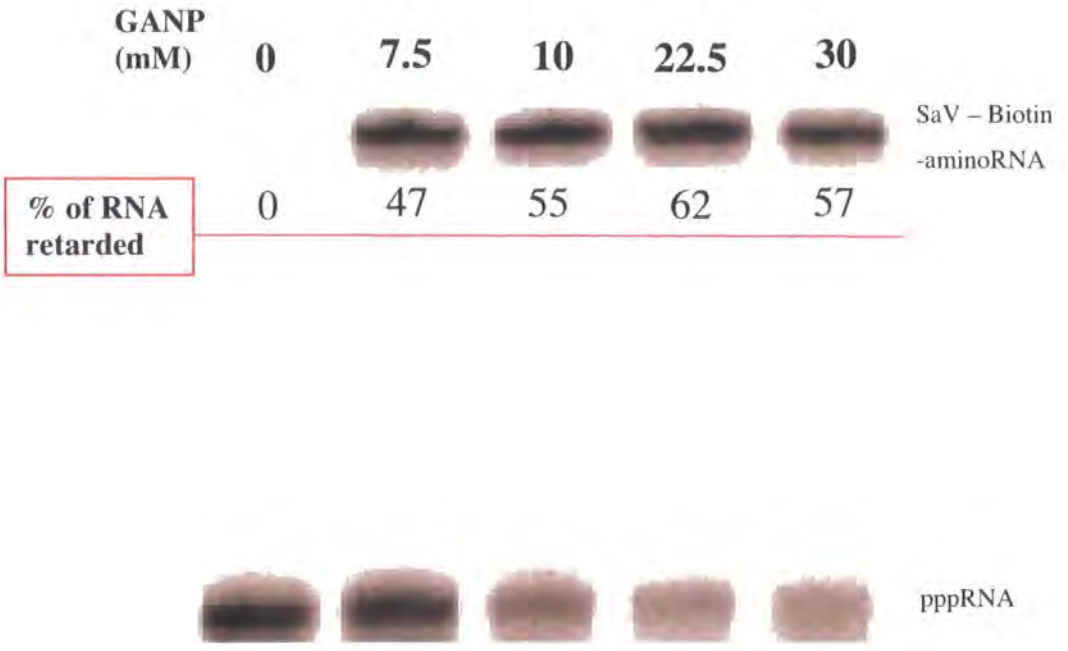
Biotinylation reactions contained 67 mM HEPES (pH 8.1) using 3.3 mM sulfo-NHS-Biotin (Sigma) and approx. 1 μM RNA. Each reaction mixture was incubated at room temperature for 2h. Reactions were stopped *via* the addition of 3 volumes of ethanol, which precipitated the RNA species. The RNA precipitate was collected by centrifugation. The pellets were washed with ethanol–water (7:3, 30 μL) before being dissolved in DEPC water (2 μL). The RNA was diluted 1:4 with streptavidin loading buffer (a 1:1 mixture of streptavidin solution [2 mg/mL]: 8M urea loading buffer) and analysed on a thin (0.4 mm thickness) urea-PAGE gel (1.5 h, 8W). RNA bands were imaged using an IP screen and Fujifilm FLA 3000 phosphorimager, and the relative intensities of retarded and non-retarded bands were used to determine relative levels of novel nucleotide incorporation into RNA transcripts.

9.24 Optimisation of Crude GANP concentration in transcription reactions[3]

Transcription reactions were carried out following the method detailed above (9.18). Following dissolution of the reaction products a sample of each was biotinylated and a streptavidin assay carried out as in the method detailed above. The table below shows average results observed when using a range of crude GANP concentration.



IP image of PAGE gel showing separation of ^{32}P incorporating reaction products



IP image of PAGE streptavidin assay

Crude GANP (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
7.5	111	47
15	93	54
22.5	65	62
30	21	56

9.25 Transcription Reactions incorporating purified GANP[3]

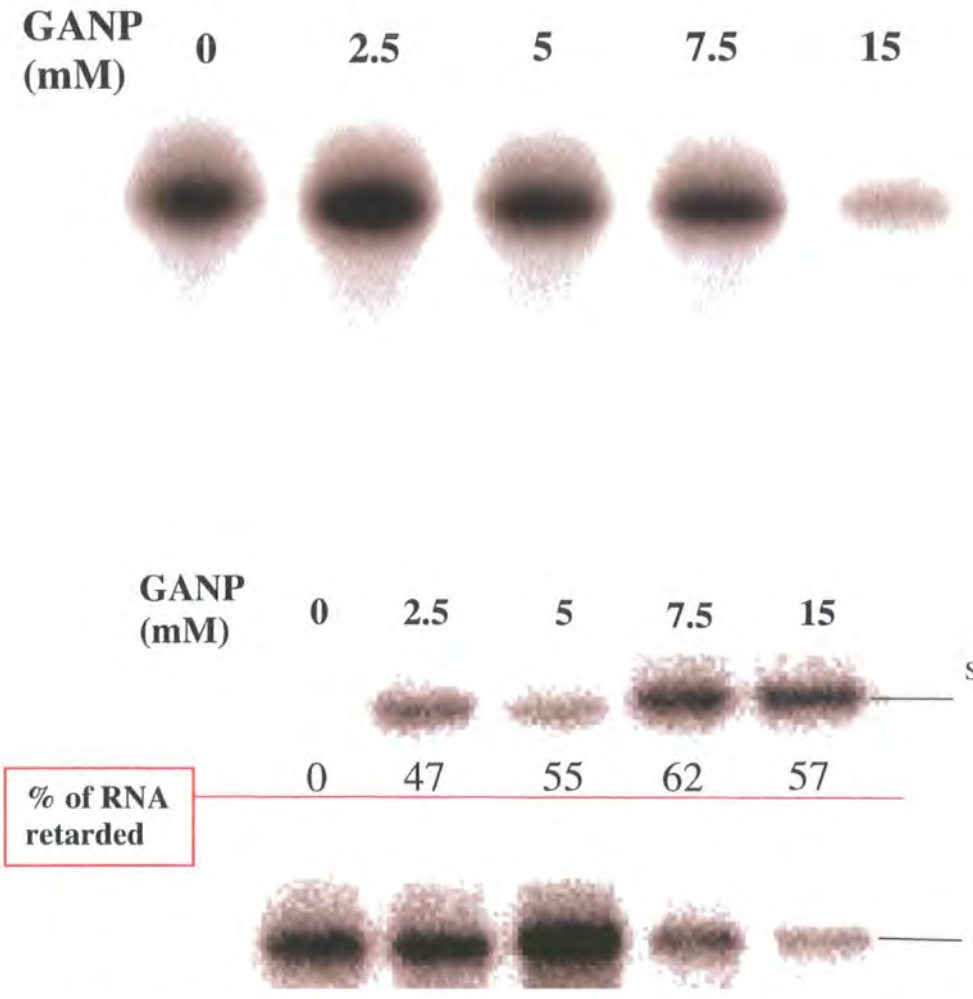
All transcription reactions using GANP purified by the method detailed in sections 2.41-57 were carried out using the protocol set out below. The only variable between reactions was the amount of GANP used in each experiment. Reagents other than GANP and T7 RNAP were combined as a stock solution before distribution between each reaction vessel in use. In this way the amount of NTPs, [α-^{32}P]-UTP and other reagents in each reaction could be standardised across the experimental range.

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase

(by UV, $\varepsilon_{280} = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$)[61]. Purified GANP was added to the reaction mixture as a 100 mM stock and diluted appropriately with DEPC water to give the desired final concentration. Reaction mixtures were incubated at 37 °C for 2.5 h. The DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. Unincoporated GANP and other small molecules were removed *via* gel filtration using a spin-column (ZEBA Desalt 0.5 ml, Pierce). The material eluted from the column was then mixed with loading buffer (1 volume, 8 M urea in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (70 % ethanol) then redissolving in DEPC water.

9.26 Optimisation of purified GANP in Transcription reactions[3]

Transcription reactions were carried out following the method detailed above (9.18). Following dissolution of the reaction products a sample of each was biotinylated and a streptavidin assay carried out as in the method detailed above. The table below shows average results gained when using a range of GANP concentrations.



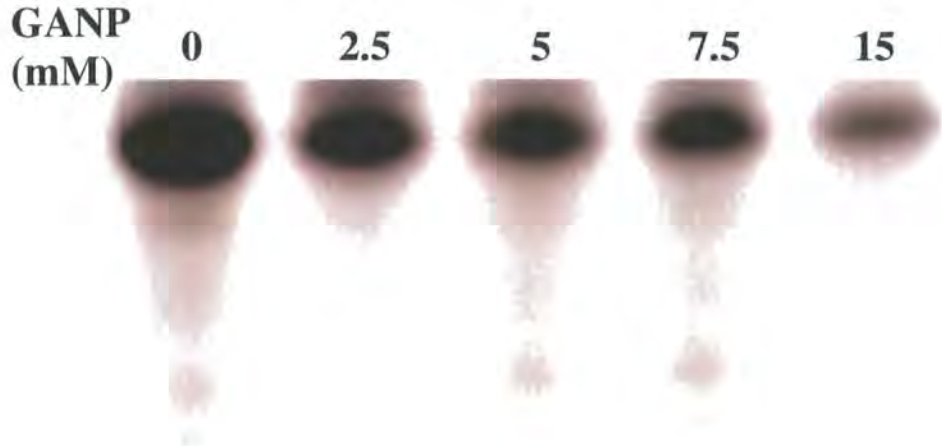
GANP (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
2.5	108	28
5	78	46
7.5	62	60
15	9	67

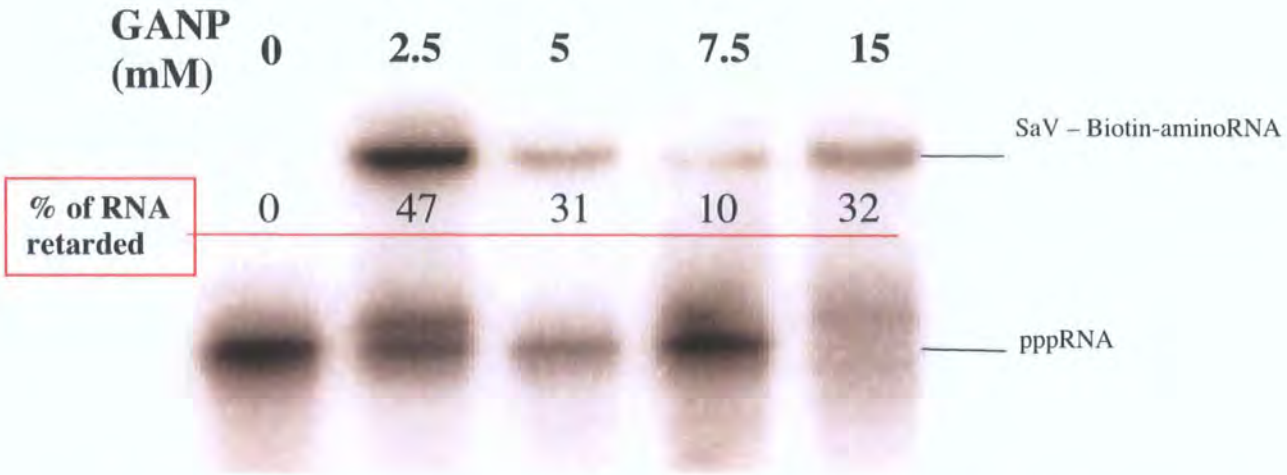
9.27 Transcription with GANP and reduced GTP concentration[3]

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of ATP, CTP, and UTP, 0.313mM GTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5$ M^{-1} cm^{-1})[61]. Purified GANP was

added to the reaction mixture as a 100 mM stock and diluted appropriately with DEPC water to give the desired final concentration. Reaction mixtures were incubated at 37 °C for 2.5 h. The DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. Unincoporated GANP and other small molecules were removed *via* gel filtration using a spin-column (ZEBA Desalt 0.5 ml, Pierce). The material eluted from the column was then mixed with loading buffer (1 volume, 8 M urea in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (70 % ethanol) then redissolving in DEPC water.

A sample of the RNA isolated from each experiment was biotinylated and used to run a streptavidin assay as described in section 9.22.





GANP (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
2.5	70	47
5	60	31
7.5	25	10
15	2	32

9.28 Transcription Reactions incorporating Amino–G[3]

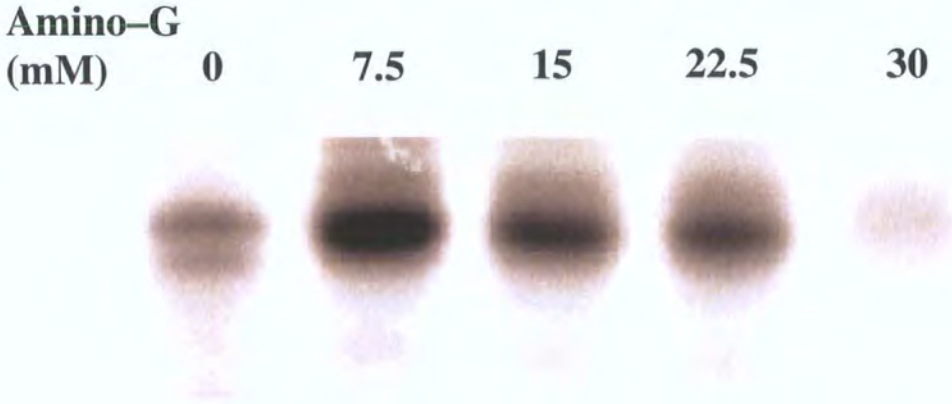
The majority of transcription reactions using Amino–G as synthesised and purified by the method detailed in section 2.32 were carried out using the protocol set out below (the exception being reactions with reduced GTP conc, see section 5.14). The amount of Amino–G was varied between reactions. Reagents other than Amino–G and T7 RNAP were combined as a stock solution before distribution between each reaction vessel in use. In this way the amount of NTPs, [α-^{32}P]-UTP and other reagents in each reaction could be standardised across the experimental range.

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$)[61]. Amino–G was added to the reaction mixture as a stock (100 mM in 100mM $NaOH_{(aq)}$) and diluted appropriately with DEPC water to give the desired final concentration. Reaction mixtures were incubated at 37 °C for 2.5 h. The

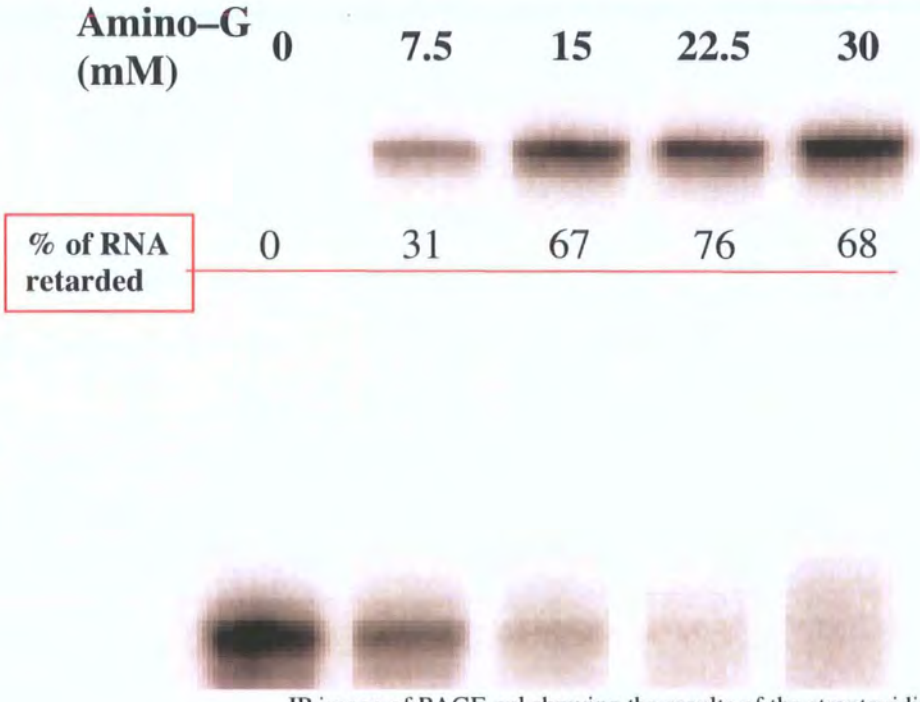
DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. Unincoporated Amino–G and other small molecules were removed *via* gel filtration using a spin-column (ZEBA Desalt 0.5 ml, Pierce). The material eluted from the column was then mixed with loading buffer (1 volume, 8 M urea in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (70 % ethanol) then redissolving in DEPC water.

9.29 Optimisation of Amino–G concentration in transcription reactions[3]

Transcription reactions were carried out following the method detailed above (9.18). Following dissolution of the reaction products a sample of each was biotinylated and a streptavidin assay carried out as in the method detailed above. The table below shows average results observed when using a range of Amino–G concentrations.



IP image of PAGE gel showing separation of transcription products.



IP image of PAGE gel showing the results of the streptavidin assay.

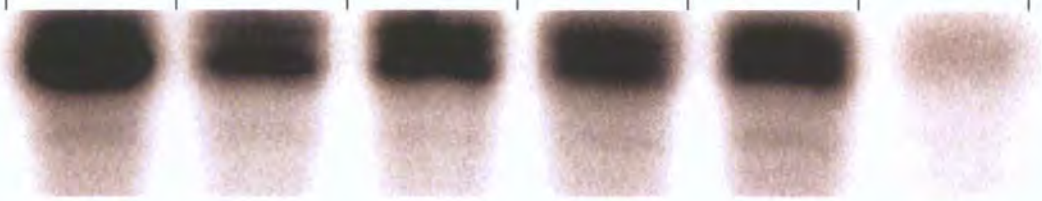
Amino–G (mM)	Total RNA yield (%)	% AminoRNA
0	100	0
7.5	139	31
15	102	67
22.5	80	76
30	25	68

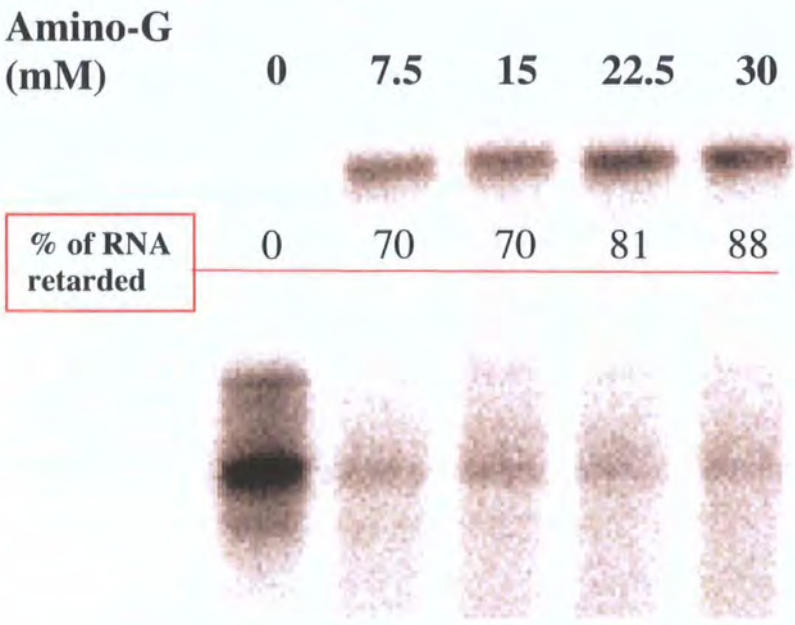
9.30 Transcription with Amino–G and reduced GTP[3]

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of ATP, CTP, and UTP, 0.313mM GTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$)[61]. Amino–G was added to the reaction mixture as a stock (100 mM in 100mM $NaOH_{(aq)}$) and diluted appropriately with DEPC water to give the desired final concentration. Reaction mixtures were incubated at 37 °C for 2.5 h. The DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. Unincoporated Amino–G and other small molecules were removed *via* gel filtration using a spin-column (ZEBA Desalt 0.5 ml, Pierce). The material eluted from the column was then mixed with loading buffer (1 volume, 8 M urea in DEPC water) and heated at 95 °C for 1 min. The resulting solution was then immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (70 % ethanol) then redissolving in DEPC water.

A sample of the RNA isolated from each experiment was used to run a streptavidin assay as described in section 9.22.

GTP (mM)	1.25	0.3	0.3	0.3	0.3	0.3
Amino-G (mM)	0	0	7.5	15	22.5	30





Amino-G (mM)	Total RNA yield (%)	% AminoRNA
Control	100	0
0	46	0
6	54	70
12	50	70
18	56	81
24	10	88

9.31 Transcription Reactions incorporating Azido–G[3]

The majority of transcription reactions using Azido–G as synthesised and purified by the method detailed in Section 2.30 were carried out using the protocol set out below. The only variable between reactions was the amount of Azido–G used in each experiment. Reagents other than Azido–G and T7 RNAP were combined as a stock solution before

distribution between each reaction vessel in use. In this way the amount of NTPs, [α-^{32}P]-UTP and other reagents in each reaction could be standardised across the experimental range.

Transcription reactions contained 40 mM Tris (pH 7.9), 10 mM DTT, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine, 350-500 nM dsDNA template*, 1.25 mM of each NTP, 83 nM [α-^{32}P]-UTP (stock contained 10 μCi/μL, 800 Ci/mM), 420 nM T7 RNA Polymerase (by UV, $\varepsilon_{280} = 1.4 \times 10^5$ M^{-1} cm^{-1})[61]. Azido–G was added to the reaction mixture as a stock (100 mM in 100mM $NaOH_{(aq)}$) and diluted appropriately with DEPC water to give the desired final concentration. Reaction mixtures were incubated at 37 °C for 2.5 h. The DNA template was then degraded through addition of RQ1 RNase-free DNase (1U, Promega) and incubation at 37 °C for a further 0.5 h. The resulting solution was immediately loaded onto a preparative (1.5 mm thickness) denaturing (urea) polyacrylamide gel (12 % cross–linking). Products were separated by electrophoresis (25 W, 2 h), the relative amounts of RNA produced in each transcription reaction were determined by exposing the gel to an IP screen and reading the screen using a Fujifilm FLA 3000 phosphorimager. The RNA bands were then located by UV shadowing, the bands were excised and the RNA was then extracted from the gel by passive elution overnight into 0.3 M NaCl on an end-over-end rotator. Gel pieces were removed using a filtration spin column (0.5 ml, Pierce). The RNA transcripts were precipitated from the eluted solution by addition of 3 volumes of ethanol and cooling to –20 °C. The solid was isolated by centrifugation before washing with ethanol solution (70 % ethanol) then redissolving in DEPC water.

9.32 5′–AzidoRNA reduction and biotinylation

Azide reductions were carried out by incubating 5′–AzidoRNA isolated from a transcription reaction in a HEPES (200 mM) buffered aqueous solution containing Tris(2-carboxyethyl)phosphine hydrochloride (100 mM) for one hour. Reaction products were precipitated through the addition of ethanol (2.2 volumes) and cooling to –20 °C. The RNA solids were then isolated through centrifugation and washed with ethanol–water (7:3, 30 μL).

Biotinylation reactions contained HEPES buffer (67 mM, pH 8.1) using sulfo-NHS-Biotin (3.3 mM, Sigma) and RNA (approx. 1 μM mixture of pppRNA and AzidoRNA). Each reaction mixture was incubated at room temperature for 2h. Reactions were stopped *via* the addition of 3 volumes of ethanol, which precipitated the RNA species. The solid RNA was collected by centrifugation. The pellets were washed with ethanol–water (7:3, 30 μL) before being dissolved in DEPC water (2 μL). The RNA was diluted 1:4 with streptavidin loading buffer (a 1:1 mixture of streptavidin solution [2 mg/mL]: 8M urea loading buffer) and analysed on a thin (0.4 mm thickness) urea-PAGE gel (1.5 h, 8W). RNA bands were imaged using an IP screen and Fujifilm FLA 3000 phosphorimager, and the relative intensities of retarded and non-retarded bands were used to determine relative levels of incorporation novel nucleotide incorporation into RNA transcripts.

9.33 Direct coupling to 5′–AzidoRNA[64]

Biotinylation reactions contained copper (II) sulfate (0.005 M), sodium ascorbate (0.023 M), alkyne–LC–biotin (0.234 M) and RNA (approx 1 μM mixture of pppRNA and AzidoRNA). The reaction mixture was incubated for 48 h under $N_{2(g)}$ during which time an orange/brown solid precipitated. Reactions were stopped *via* the addition of 3 volumes of ethanol, which precipitated the RNA species. The solid RNA and orange/brown precipitate was collected by centrifugation. The pellets were washed with ethanol–water (7:3, 30 μL) before being suspended in DEPC water (2 μL). The RNA was diluted 1:4 with streptavidin loading buffer (a 1:1 mixture of streptavidin solution [2 mg/mL]: 8M urea loading buffer) and analysed on a thin (0.4 mm thickness) urea-PAGE gel (1.5 h, 8W). RNA bands were imaged using an IP screen and Fujifilm FLA 3000 phosphorimager.

(1) Jaschke, A.; Seelig, B. *Current Opinion in Structural Biology* **2000**, *4*, 257-262.
(2) Voet, D. V., J.G. *Biochemistry 3rd Ed.*; John Wiley and sons inc.
(3) Milligan, J. F. *Nucleic Acids Research* **1987**, *15*, 8783-8798.
(4) Cheetham, G. M., .T; Jeruzalmi, D.; Steitz, T. A. *Nature* **1999**, *399*, 80-83.
(5) Cheetham, G. M., .T; Steitz, T. A. *Science* **1999**, *286*, 2305-2309.
(6) Milligan, J. F.; Uhlenbeck, O. C. *Methods in Enzymology* **1989**, *180*.
(7) Steitz, T. A. *Science* **1994**, *266*, 2022-2026.
(8) Cazenave, C.; Uhlenbeck, O. C. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6972-6976.
(9) Wecker, M.; Smith, D.; Gold, L. *RNA* **1996**, *2*, 982-994.
(10) Jaschke, A.; Friedrich, S. *Journal of the American Chemical Society* **2002**, *124*, 3238-3244.
(11) Jaschke, A.; Hausch, F. *Nucleic Acids Research* **2000**, *28*, e35.
(12) Jaschke, A.; Seelig, B. *Tetrahedron Letters* **1997**, *38*, 7729-7732.
(13) Szostak, J. W.; Suga, H.; Lohose, P. A. *Journal of the American Chemical Society* **1998**, *120*, 1151-1156.
(14) Eckstein, F.; Aurup, H.; Williams, D. M. *Biochemistry* **1992**, *31*, 9636-9641.
(15) Sousa, R.; Brieba, L. G.; Padilla, R. *Biochemistry* **2002**, *41*, 5144-5149.
(16) Huang, F. Y., Changjun. L, Na *Nucleic Acids Research* **2005**, *33*, e37.
(17) Cheetham, G. M., .T; Steitz, T. A.; Jeruzalmi, D. *Nature* **1999**, *399*, 80-83.
(18) Sousa, R.; Brieba, L. G. *Biochemistry* **2000**, *39*, 919-923.
(19) Martin, C. T. *The Journal of Biological Chemistry* **2003**, *278*, 2819-2823.
(20) Coleman, J. E.; Martin, C. T. *Biochemistry* **1989**, *28*, 2760-2762.
(21) Jaschke, A.; Hausch, F. *Tetrahedron* **2001**, *57*, 1261-1268.
(22) Cech, T. R.; Zhang, B. *Nature* **1997**, *390*, 96-100.
(23) Pan, T. *Current Opinion in Chemical Biology* **1997**, *1*, 17-25.
(24) Jaschke, A.; Frauendorf, C. *Bioorganic and Medicinal Chemistry* **2001**, *9*, 2521-2524.
(25) Zhang, B.; Zhang, L.; Cui, Z.; Gottlieb, R. L. **2001**.
(26) Zhang, B.; Cui, Z.; Sun, L. *Organic Letters* **2001**, *3*, 275-278.
(27) Famulok, M.; Jaschke, A.; Seelig, B.; Sengle, G. *Bioorganic and Medicinal Chemistry* **2000**, *8*, 1317-1329.
28) Zhang, B.; Zhang, L.; Sun, L.; Cui, Z.; R.L., G. *Bioconjugate Chemistry* **2001**, *12*, 939-948.
(29) Zhang, B.; Sun, L.; Cui, Z.; Gottlieb, R. L. *Chemistry and Biology* **2002**, *9*, 619-628.
(30) Burgin, A. P., N. *The EMBO Journal* **1990**, *9*, 4111-4118.
(31) Jaschke, A. S., J. *Biochemical and Biophysical Research Communications* **2006**, *344*, 887-892.
(32) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press, 1996.
(33) Debye *Phys.Colloid.Chem* **1947**, *51*, 18.
(34) *Molecular Probes Catalogue* **2007**.
(35) Diagnostics, K.: EP Patent. 0539 466 B1
(36) Nakane, P. K. K., A. *J. Histochem. Cytochem* **1974**, *22*, 1084-1091.
(37) Keller, G. H. M., M.M *Analytical Biochemistry* **1988**, *170*, 441-450.
(38) Kricka, L. J. *Nonisotopic DNA Probe Techniques*; Academic Press, New York, 1992.
(39) Chu, B. C. F. O., L.E *Nucleic Acids Research* **1983**, *11*, 5591-5603.
(40) Chu, B. C. F.; Kramer, F. R.; Orgel, L. E. *Nucleic Acids Research* **1986**, *14*, 5591-5603.
(41) Ghosh *Bioconjugate chemistry* **1990**, *1*, 71-76.
(42) Chu, B. C. F.; Orgel, L. E. *Nucleic Acids Research* **1988**, *16*, 3671-3691.
(43) He, B. R., M.Q. Lyakov, D. *Protein Expression and Purification* **1997**, *9*.
(44) Jaschke, A. F., R. Musilek, K. *Journal of the American Chemical Society* **2005**, *127*, 9271-9276.
(45) Suga, H.; Cowan, J. A.; Szostak, J. W. *Biochemistry* **1998**, *37*, 10118-10125.
(46) Dean, D. K. *Synthetic Communications* **2002**, *32*, 1517-1521.
(47) Hampton, A.; Brox, L. W.; Bayer, M. *Biochemistry* **1969**, *8*, 2303-2311.
(48) Chanley, J. D.; Feageson, E. *Journal of the American Chemical Society* **1958**, *80*, 2686.
(49) Chanley, J. D.; Feageson, E. **1963**, *85*, 1181-1190.

(50) Benkovic, S. J., Sampson, E.J *Journal of the American Chemical Society* **1971**, *93*, 4009-4016.
(51) Jencks, W. P. H., D *Journal of the American Chemical Society* **1989**, *111*, 7579 – 7586.
(52) Jastorff, B.; Hettler, H. *Chem. Ber.* **1969**, *102*, 4119-4127.
(53) Jastorff, B.; Hettler, H. *Tetrahedron Letters* **1969**, *30*, 2543-2544.
(54) McGee, D. P. C.; Martin, J. C. *Canadian Journal of Chemistry* **1986**, *64*, 1185-1189.
(55) Yoshikawa, M.; Kato, T.; Takenishi, T. *Bulletin of the Chemical Society of Japan* **1969**, *42*, 3505-3508.
(56) Yoshikawa, M.; Kato, T.; Takenishi, T. *Tetrahedron Letters* **1967**, *50*, 5065-5068.
(57) Whitesides, G. M.; Crans, D. C. *Journal of the American Chemical Society* **1985**, *107*, 7008-7018.
(58) Drueckhammer, D. G.; Duncan, R. *Tetrahedron Letters* **1993**, *34*, 1733-1736.
(59) Berger, S. S., B. Kalinowski, H.O. *NMR spectroscopy of the non-metallic elements*; Wiley. Chichester, 1997.
(60) *Promega Catalogue*, 2006.
(61) He, B. R., M.Q. Lyakov, D. *Protein Expression and Purification* **1996**, 142-151.
(62) FUJIFILM *www.fujifilm.com/products/lifesciences/si_imgplate/img_plate.html*.
(63) Chamberlin, M. R., J. *The journal of biological chemistry* **1973**, *248*, 2235-2244.
(64) Schuber, F. F., B. Fatouma, S.H. *Bioconjugate Chemistry* **2006**, *17*, 849-854.
(65) Bertozzi, C. R. S., E. *Science* **2000**, *287*, 2007-2010.
(66) Yamamoto, Y. K., S. *Tetrahedron letters* **2004**, *45*, 689-691.
(67) Brown, T. K., R. El–Sagheer, A. *The journal of the american chemical society* **2007**, *129*, 6859-6884.
(68) Chan, T. R. H., R. Sharpless, K.B. *Organic letters* **2004**, *16*, 2853-2855.